PROCEEDINGS

OF THE

AMERICAN PHILOSOPHICAL SOCIETY

HELD AT PHILADELPHIA
FOR PROMOTING USEFUL KNOWLEDGE

VOLUME 88

1944

THE AMERICAN PHILOSOPHICAL SOCIETY
INDEPENDENCE SQUARE
PHILADELPHIA 6

1944

R506 AC
v.88 June-Dec. 1944 c.1

COMMITTEE ON PUBLICATIONS

JACOB R. SCHRAMM, *Chairman*

LUTHER P. EISENHART
WILLIAM K. GREGORY
HENRY C. LANCASTER
WILLIAM E. LINGELBACH
FOREST R. MOULTON
ARTHUR D. NOCK
CONYERS READ
ADOLPH H. SCHULTZ
ROBERT L. SCHUYLER
T. LESLIE SHEAR

LUTHER P. EISENHART, *Editor*

GCI

LANCASTER PRESS, INC., LANCASTER, PA.

CONTENTS

Page

No. 1. Symposium on Taxation and the Social Structure
- The Capitalist and the Social Structure. ROBERT WARREN ... 1
- Taxes and National Income. SIMON KUZNETS ... 10
- Conflict and Harmony in Taxation. ROY BLOUGH ... 22
- The Sales Tax. ALFRED G. BUEHLER ... 29
- Fiscal Policy and Social Reform as it May Affect the Potentialities of the Personal Income Tax. HENRY SIMON BLOCH ... 36
- The Harmonization of Fiscal and Social Aims. MERLIN H. HUNTER ... 43
- The Federal, State, and Local Tax Structure after the War. MABEL NEWCOMER ... 50
- Post-War Social Orientation of State Tax Systems. WILLIAM J. SHULTZ ... 55
- A Post-War Program for the Taxation of Corporations and their Stockholders. EUSTACE SELIGMAN ... 60

No. 2. Catawba Pottery-making, with Notes on Pamunkey Pottery-making, Cherokee Pottery-making, and Coiling. VLADIMIR J. FEWKES ... 69
- Recovery of Midwestern Prairies from Drought. JOHN E. WEAVER ... 125
- Some Viruses Transmitted by Agallian Leafhoppers. L. M. BLACK ... 132
- Some Tables to Facilitate Computation of Elements of Eclipsing Binaries. ZDENĚK KOPAL 145

No. 3. Symposium on Wartime Advances in Medicine
- The Water Requirements of Castaways. JAMES L. GAMBLE ... 151
- Blood, Blood Derivatives, and Blood Substitutes. EDWIN J. COHN ... 159
- The Present Status of Penicillin in the Treatment of Infections. CHESTER S. KEEFER ... 174
- Recent Advances in the Treatment of Wounds. ALLEN O. WHIPPLE ... 177
- The Role of Immunization in Wartime Advances in Medicine. R. E. DYER ... 182
- Human Problems in Military Aviation. DETLEV W. BRONK ... 189
- Environmental Protection. GEORGES F. DORIOT ... 196
- Neuropsychiatry. EDWARD A. STRECKER ... 204
- Trends in the Study and Control of Infectious Diseases. RENÉ J. DUBOS ... 208

R. A. F. Penrose, Jr., Memorial Lecture
- The International Labor Organization Lessons for Twenty-five Years. HAROLD B. BUTLER 214

No. 4. Mountain Growth, a Study of the Southwestern Pacific Region. WILLIAM HERBERT HOBBS ... 221
- Figural After-effects. An Investigation of Visual Processes. WOLFGANG KÖHLER AND HANS WALLACH ... 269

No. 5. "Of the People, By the People, For the People." JOHN L. HANEY ... 359
- Dakota Treatment of Murderers. ELLA CARA DELORIA ... 368
- A Photographic Study of the Proper Motion of Nearby Stars. PETER VAN DE KAMP ... 372
- Postnatal Development of the Human Extremities. CHARLES BENEDICT DAVENPORT ... 375

No. 6. Lagrange's Personality. GEORGE SARTON ... 457
- The Life Portraits of Jefferson and their Replicas. FISKE KIMBALL ... 497
- Index ... 535

Periodical

PUBLIC LIBRARY
JUN 30 1944
DETROIT

PROCEEDINGS

OF THE

American Philosophical Society

Held at Philadelphia for Promoting Useful Knowledge

VOLUME 88, NUMBER 1 JUNE 16, 1944

SYMPOSIUM

ON

TAXATION AND THE SOCIAL STRUCTURE

Papers read before
The American Philosophical Society
Midwinter Meeting
February 18–19, 1944

PHILADELPHIA

THE AMERICAN PHILOSOPHICAL SOCIETY

INDEPENDENCE SQUARE

1944

COMMITTEE ON PUBLICATIONS

JACOB R. SCHRAMM, *Chairman*

LUTHER P. EISENHART	ARTHUR D. NOCK
WILLIAM K. GREGORY	CONYERS READ
HENRY C. LANCASTER	ADOLPH H. SCHULTZ
WILLIAM E. LINGELBACH	ROBERT L. SCHUYLER
FOREST R. MOULTON	T. LESLIE SHEAR

LUTHER P. EISENHART, *Editor*

LANCASTER PRESS, INC., LANCASTER, PA.

SYMPOSIUM

on

TAXATION AND THE SOCIAL STRUCTURE

Papers read before

The American Philosophical Society

Midwinter Meeting

February 18–19, 1944

PHILADELPHIA

THE AMERICAN PHILOSOPHICAL SOCIETY

INDEPENDENCE SQUARE

1944

CONTENTS

The Capitalist and the Social Structure 1
ROBERT WARREN, Institute for Advanced Study.

Taxes and National Income 10
SIMON KUZNETS, University of Pennsylvania.

Conflict and Harmony in Taxation 22
ROY BLOUGH, United States Treasury Department.

The Sales Tax 29
ALFRED G. BUEHLER, University of Pennsylvania.

Fiscal Policy and Social Reform as it may affect the Potentialities of the Personal Income Tax 36
HENRY SIMON BLOCH, University of Chicago.

The Harmonization of Fiscal and Social Aims 43
MERLIN H. HUNTER, University of Illinois.

The Federal, State, and Local Tax Structure after the War 50
MABEL NEWCOMER, Vassar College.

Post-War Social Orientation of State Tax Systems 55
WILLIAM J. SHULTZ, College of the City of New York.

A Post-War Program for the Taxation of Corporations and their Stockholders ... 60
EUSTACE SELIGMAN, New York.

THE CAPITALIST AND THE SOCIAL STRUCTURE

ROBERT WARREN

Institute for Advanced Study

(Read February 18, 1944, in the Symposium on Taxation and the Social Structure)

As I see it, political economy is the study of the problem of the reconciliation of social aspirations with the state of the arts.

The term "state of the arts" is borrowed from Thorstein Veblen, who meant the technology of the time in its institutional or functional setting. In the course of the nineteenth century the state of the arts developed a distinct form which came to be called capitalism; and the actuality was well advanced before the word gained currency in the United States. Today the term is commonly used in an antagonistic rather than a descriptive sense, and its adherents prefer nondescript circumlocutions, such as the enterprise system, or the American way. As a result of what might be called the Academic Lag, there are no such objective descriptions of nineteenth century capitalism, as there are of mercantilism or feudalism.

By "social aspirations" I mean the rather vague but deeply held notions of the nature of the "good society." There is a folklore of the "good society"; it is largely expressed in familiar quotations from the Founding Fathers; like all folklore, it has its mythical heroes who have become personifications of ideals; and it finds expression in recurrent political movements, and in legislation.

Capitalism is the state of the arts that evolved from the collision between the industrial revolution and the money economy, superseding the mercantilism that had evolved from the collision between the money economy and feudalism. Technically, capitalism brought an immense increase in the complexity, power, and cost of the instruments of production. Socially, it involved a degree of collectivism beyond anything the world had previously seen. Institutionally, the characteristic collective of capitalism was the corporation.

As capitalism developed, society itself was undergoing a structural change, partly in consequence of a change in scale. As far as one can observe, in each nation the arrival of the industrial revolution has been promptly followed by a fall in the deathrate, and a rapid increase in the population. This phase is accompanied by urbanization, and this, in turn, by an eventual slow decline in the birthrate. The social effects of these changes may be enumerated as follows:

1. Urbanization.
2. Two changes in the age composition of the population—an initial period when the number of young is relatively large (because of the decline in infant mortality) and a subsequent period when the number of young is relatively small (because of the fall in the birthrate).

These changes in scale and composition were not accompanied *pari passu* by changes in our habits of thought. For the first time in history we have a society without a common traditional wisdom. A society undergoing such changes will not lose, but it will modify its aspirations; and it will find no little difficulty in making the changes articulate. Even common words change their content without losing their currency. When Ethan Allen's turbulent Green Mountain Boys marched to Ticonderoga under a banner inscribed "Liberty and Property," they did not dream of a day when "property rights" would be called hostile to "human rights." Property has a different connotation to an urban, industrialized society than to rural, agrarian society.

As a result of its rapid occurrence, there is no folklore of capitalism. Our social folklore dates from the pre-capitalistic days, and its cherished phrases regarding self-evident truths, the consent of the governed, and the rights of man, which rang and still ring so clearly in the ears of an agrarian society, are but the notes of an uncertain trumpet to an urbanized and industrialized community.

In short, while we speak rather glibly of social aspirations, of authentic spokesmen of the national genius, inherent rights, and so forth, we are dealing with extremely difficult and elusive concepts. We are dealing with ideas that move over time and space, rather than with constant and universal truths. Yet with all the change,

there are continuities of social aspiration, even though they appear in new forms and dressed in novel nomenclatures. The specific problem changes, but the problem itself never changes. The problem of political economy is always to reconcile social aspirations to the state of the arts.

Tonight I propose to discuss a single aspect of this problem: namely, the relation of a character called the capitalist to that state of the arts called capitalism, and the relation of the capitalist to certain social aspirations as expressed in our tax legislation.

By definition the capitalist is a rich or relatively rich man. That is, he either commands a fortune of some magnitude, or an income well above the demands of his personal consumption. These facts distinguish him from the great majority of mankind. But, to limit the discussion, not all rich men and not all men with large incomes are capitalists. There is also a difference between capitalists and financiers, who are merchants of capital. The capitalist of tonight's discussion is the Marxian capitalist, who owned the tools of production used by others and who took his profit out of what Marx termed "surplus value" and at times, at least, out of what Henry George called "unearned increment"; and who took his losses out of his own pocket.

The capitalist of tonight's discussion is the species of fauna *dives anthropophagus* as distinct from *dives gaudens*, celebrated by the gossip columns; or *dives otiosus*, the hero of Veblen's best-known book; or *dives philanthropus*, the patron saint of learning and healing.

The capitalist, then, is an example of economic inequality. Economic inequality is as old as society itself, and society has always been under some constraint to rationalize it. Under feudalism, economic inequality was almost invariably a derivative of inherited status, and was explained by ramifications of the theory of divine right. Under mercantilism, the novelty of uninherited wealth made its appearance; and as Professor Tawney elaborates, it was explained in terms of specific divine grace.

> The rich man in his castle,
> The poor man at his gate,
> God made them high and lowly,
> And ordered their estate.

This idea, while obviously of wider application than the doctrine of hereditary right, weakened before the more pragmatic concept of the self-made man. This term had and still has genuine social validity, although it is impaired by recognition of the fact that, as Henry George pointed out, the self-made man is to a considerable degree the product of his neighbors. Nevertheless, wealth as such—that is, wealth not more or less definitely associated with personal achievement—has never in this country been accorded anything like the deference rendered in many societies, and we may imagine that this American attitude is one of the legacies of the egalitarianism of the frontier, so much emphasized by Professor Turner. The formula accepted in Soviet Russia to explain economic inequality—"From each according to his ability, to each according to his contribution" is, as formulas go, not too bad an expression of American attitude, except that it belittles the difficulty of appraising the value of the individual's contribution under the complicated division of labor characteristic of either state capitalism or private capitalism. The utility of this formula is that it reduces economic inequality to terms more manageable than divine grace and permits us to discuss the capitalist in purely functional terms.

What contribution, if any, did the capitalist make to society? What was his functional relation to the state of the arts?

As I see it, the capitalist performed three recognizable functions:

1. The function of entrepreneurship.
2. The function of representing the proprietary interest in corporate enterprise.
3. The function of combating monopoly.

The entrepreneur—the man who initiates new enterprises—was a dealer in risk. He was not a gambler, for the risks were not aleatory. He was not an investor, for an investor places his money in a going concern without assuming managerial responsibility. The capitalist might, and of necessity usually did, contribute his energy as well as his money to the undertaking; and in the nature of things operated only with his own funds, since the hazard of new ventures exceeded the bounds of fiduciary relation. Acting under the assurance that any given undertaking might, and that of a given number of undertakings, a large proportion certainly would, result in total loss, he calculated his return not in a rate percent per annum, but as an equation of loss and gain over an uncertain period of time.

Entrepreneurship has become continuously

more expensive. As we said earlier, the technological characteristic of capitalism has been the progressive increase in the cost of the tools of production, or equipment per worker. In the days of '49, many a profitable placer was opened with no more capital than was represented by a burro, a pick, a shovel, and washing pan, and a six months' grubstake for the proprietor. In Canada, today, a gold prospect may expend $500,000 before the first ton of ore is brought to the surface. In recent years a new venture, contemplating the employment of other than the owners' own services, could expect an initial outlay of from $2,000 to $20,000 per person prospectively employed, plus the costs of an uncertain number of years of unprofitable operations—in the hope that in the end it would pay out. But only the hope. I would venture the guess that in mining, out of 100 prospects, each one representing a considerable amount of preliminary expenditure, ten might justify more or less costly development work; and of these ten, one might or might not turn into a mine. There is no reason to suppose that mining is peculiar or unique in the area of new enterprise.

In pre-capitalistic days, the new-fledged journeyman received from his master the tools of his trade. These represented a kind of capital accumulation from the unpaid services of apprenticeship; and so equipped, the young man fared forth to seek his fortune. Their entire value might be a few shillings or a few pounds at most. Today, in many kinds of industry, as far as one can make out from rather inadequate statistics, the fixed capital investment per worker will run to $10,000, $20,000, or even higher. Even in agriculture, which we commonly think of as a non-capitalistic enterprise, the capital investment in buildings, equipment, and livestock —exclusive of land—will commonly run $4,000 or $5,000 per adult male farm worker. In the present state of the arts, the equivalent of the tools of his trade given to the passed apprentice is equipment costing several or many thousand dollars, and this expenditure must have been made and must have been installed before what we call a "job" exists. It must be there the day before the new recruit to the labor force applies for his first job. Under contemporary capitalism, the tools of the trade are not only very expensive, but they are useless except as they are integrated with the tools of a more or less numerous group. In short, venture capital, while a variable, long since became a variable of rapidly rising magnitude; and the provision of venture capital the function of a progressively richer capitalist.

The second function of the capitalist, as I see it, was to represent the proprietary interest in the organization of corporate capitalism. This is a highly disputed topic, and for that reason I wish to emphasize it. We have spoken of the corporation as a complex collectivism. This collectivism is composed of the so-called employees, the managerial hierarchy, the legal owners, and, often, creditors. The employees are organized; the managerial group is a hierarchy; the creditors have the legal protection of their bond, and are commonly a small number of powerful institutions. The so-called owners or shareholders are ostensibly organized after the pattern of a democratic government. But a succession of writers has called attention to the fact that the extreme diffusion of share ownership has radically affected the quality of this ownership function. One of the concerns of Mr. Brandeis, writing early in the century, was the fact that a degree of ownership representing but a small part of the whole could acquire great power of control over the utilization of other people's money. A generation later Messrs. Berle and Means repeated this charge with a considerable amount of documentary evidence, indicating that, by reason of this diffusion of ownership, a comparatively small number of persons controlled, or hypothetically could control, all the large corporations of America. The reason is simple. In many corporations the largest holding of any individual shareholder amounts to less than 10 percent, and often less than 2 percent, of the total stock outstanding, the remainder being scattered in minute amounts among a multiplicity of investors, of whom few or none have the slightest knowledge of the corporation, or any capacity to affect its actions. Legally they are owners, but they have none of the responsibility that we associate with the word. In short, as these analysts saw it, the diffusion of modern property facilitated a concentration of control and of economic power in the hands of a few capitalists. This Charybdis was alarming enough; but scarcely more alarming than the Scylla explored more recently by Professor Burnham in his *Managerial Revolution*. His thesis is that, while hypothetically any capitalist who could get control of a comparatively small proportion of the stock of any one of these supercorporations could control the corporation, the effective control lies in the hands of the managerial hierarchy. This often has little or no

ownership status, and is so completely independent in its conduct that, while it may not act contrary to the interests of the owners, it is under no pressure to subordinate its own interests to that of the owners in case the two conflict.

In a sense, these conclusions are poles apart; in another sense they coincide. The actual or potential concentration of economic power remains under either alternative, for it is inherent in the magnitude of the contemporary corporate unit. But from the standpoint of our argument, under one alternative the powerful stockholder (or capitalist as we have called him) represents the proprietary interest in the active direction of the corporation; and there is some presumption that his interests and those of his fellow stockholders are more or less parallel. Under the other alternative, the alleged owners may have no representative of their own near the seat of power. In so far as ownership has any social value in our scheme of things, there is at least something to be said for this ownership having an appropriate responsibility; and something to be said for the capitalist who held the position of a self-constituted representative of his legal partners. Yet, in a great corporation, the minimum fractional ownership which could claim this proprietary function would represent the investment of a relatively large fortune.

The third function is of a different order, though it represents another aspect of entrepreneurship. Scarcely had the corporation appeared before its tendency toward monopoly was recognized. The social response in America was fairly prompt and took the unique form of the anti-trust acts. Our anti-trust laws are peculiarly American; they have no counterpart in the legislation of any other country, as far as I know, and they, if nothing else, have differentiated American capitalism. They were opposed at the time of passage; they have never been without powerful opposition since; they have been fitfully and sometimes rather half-heartedly enforced; and I suppose they have been evaded. Nevertheless, I believe that, in restraining or preventing the growth of such monopolies as characterize capitalism in other countries, they have genuinely reflected one of the social aspirations of the American people. But I do not attribute the absence or relative absence of monopolies in America solely to these laws. They have found in the capitalist a vigilant and aggressive ally. That is, the existence of a capitalist, willing and able to launch a new rival concern, or to rejuvenate a moribund rival, has been one of the main safeguards against monopoly; and one of the functions of the American capitalist has been to oppose the development of monopolistic concentrations of economic power.

If you believe I have drawn too favorable a picture of the capitalist as an integral part of the state of the arts in America, I can assure you that an even more complimentary sketch is presented by Mr. Brandeis in *Other People's Money*—a book which almost cost Mr. Brandeis his appointment to the Supreme Court.

Once again, having stated what we consider the economic functions of the capitalist, we shall quote Veblen, who was, it will be agreed, a not too friendly critic: "There is one thing that can be said for capitalism. It works." If Veblen had lived a decade longer, I doubt that he would have offered this rather grudging tribute. For in very recent years it has not seemed to be working—at least not so well as it seemed to work before. Specifically, it was not providing each year such a number of new opportunities for employment as would absorb the annual recruits to the labor force plus the number of persons released from previous occupations by the gradual but supposedly continuous progress of technical efficiency. Formerly we had been able not merely to take those two in our stride, but to find places for hundreds of thousands of immigrants.

As would be expected, two explanations were advanced in succession. The first, accepted in the initial phase of the Great Depression, was that the failure was temporary; that it was like the previous cyclical episodes and would presently pass. This bright hope faded slowly; and was replaced, logically enough, by its antithesis. This is the hypothesis that the phenomenon, far from being temporary, is permanent; that the dynamo has run down. This theory has received wide attention as the doctrine of the "mature economy"; but its best literary expression to date will be found in the *Degradation of the Democratic Dogma*, by Henry Adams, published in 1919. While the phrase "mature economy" has rather fallen out of vogue, the ideas that it represents have not lost adherents. Rather, I believe that the memory of the decade between the collapse of 1929 and the beginning of the war boom in 1939 has persuaded an increasing number of persons that the phenomena of the period were not cyclic, but were indicative of serious structural defects; that these defects have not yet been adequately

discovered and analyzed; and that until they are, the measures applied may mitigate the consequences but cannot cure the causes. In short, we are still in a stage of diagnosis; and it would be my guess that before the right answer is found, a good many hypotheses will have to be explored, so to speak, "island by island." I offer a hypothesis not so mechanical as the cyclical theory, nor so metaphysical as the doctrine of the mature economy, but one purely functional. You will understand that I am not advancing this extremely simple and rather dogmatic theory as an all-embracing explanation of a "tragic era." As I said a moment ago, I am attempting a diagnosis and inviting you to join in exploring an hypothesis.

I suggest that it is worth while to see if what we call the failure of capitalism is not in considerable part attributable to circumstances which have prevented the capitalist from performing his specific functions.

In at least two respects, the characteristics of the period point rather definitely in that direction. There was a conspicuous absence of the launching of new, employment-creating enterprises; and there was enough evidence of a drift toward monopoly to occasion a Senate inquiry in 1938. It would appear that not only for the decade prior to 1939, but for an uncertain period before 1929, the capitalist was not performing his ascribed functions, and for a succession of reasons. I suggest that in the so-called New Era, the capitalist absented himself from his economic functions—that he neglected his tasks of launching new enterprises, rejuvenating old enterprises, and representing proprietary responsibilities in going concerns—all of which are hazardous and arduous tasks—for the easier and ostensibly more lucrative activities of financiering. I can offer little evidence to support this idea; but I commend it to your memories of that period to see if it seems plausible. I suggest that in the second period—from 1929 to, say, the mid-thirties, the capitalist was incapacitated from performing his functions because he had lost much of his money and all of his courage. And I suggest that from the mid-thirties to the outbreak of the European war, by which time he had presumably recovered part of his money and some of his courage, he was excluded from his functions by schedules of income taxation beyond what those functions would bear.

What taxation the functions of the capitalist would bear was not nearly equal to public superstition. Out of ten well-considered and costly ventures, one might prove a bonanza and so attract public notice. For the performance of the function, the nine failures were essential to the tenth success; but they were ignored or forgotten.

Taxation, as we said above, is one of the expressions of social attitude, and in this instance I believe that it is to be related to that group of social aspirations loosely termed economic equality. The performance of the functions of the capitalist presumed economic inequality on a more or less extreme scale. Literal economic equality has never figured in American social aspirations. The accepted aim has been to create equal opportunities to achieve unequal rewards, and there has been constant effort to rationalize the resulting economic inequality—that it should be related to the social contribution; and that it should be limited at the point where it impinges on the potential social contributions of a rival. But in the late thirties our society regarded the capitalist as socially privileged rather than economically functional—which is hardly surprising, since one aspect was far more conspicuous than the other.

(At this point, I must enter a caveat. I am discussing the functions of the capitalist in a peace economy, and the impairment of those functions by taxation. These remarks are irrelevant to the immediate situation. While our income taxes on the higher brackets are more severe, I believe, than in any other country, I am not prepared to argue that they impair any military function; indeed, I would go further and offer the opinion that a very strong case can be made for an absolute ceiling on personal income in a war economy. But that, as Rudyard Kipling used to say, is another story.)

The tax legislation of the late thirties was popularly and aptly designated as "soaking the rich." The general principle of a progressive income tax expresses a social aspiration; a schedule of progression, however, may be steeper than the functions of the capitalist can bear. As a matter of record, the schedules of progression adopted in the late thirties were not remarkably productive of revenue, and the shape of the curve of progression emphasizes the social objective of "soaking the rich"; or to put the idea more semantically, of achieving a more equitable distribution of income.

This legislation drew no distinction between the personal and functional income of the capi-

talist. The simplest and most familiar function of income is to provide for the consumption needs of the recipient and his dependents. Judged by this criterion alone, the progression schedules of the late thirties could hardly be described as intolerable. The income of the rich, after Federal taxes, was still sufficient to provide them with food, clothing, and shelter, on a scale sufficient to distinguish them from plebeians; to permit them to lay something away for old age; to make those donations to worthy causes which figure rather more heavily in conversation than in the data reported in the Statistics of Income; and even to make appropriate contributions to campaign funds. Comparisons were often made with the schedules in other countries, leading to the view that the rich American got off relatively easily. In short, during the late thirties, the income of the rich after taxes was still high enough to serve abundantly the needs of personal consumption—to answer appropriately every purpose which the vast majority of people customarily associate with income.

Now, if that were the entire story, we might congratulate ourselves that in the late thirties we had effected a reconciliation between our social aspirations and the state of the arts, for it may confidently be said that there were very few law-abiding persons in this country receiving a net income after taxes so large that there was very much danger of their becoming capitalists. Or, to put it another way, much hope of their becoming entrepreneurs, or of their being able to enter the field against prospective monopolies.

Yet, at this very time, discussion was just beginning to veer away from the explanation of our unemployment in terms of cyclical fluctuations; the hypothesis of the mature economy was being invoked to explain the absence of new enterprises, the inability of corporations to finance themselves by issues of common stock, and a senatorial committee was solemnly convoked to ponder the apparent drifts toward monopoly of corporate concentration of economic power.

In short, our tax system had paraphrased the famous remark of the French revolution, "La République n'a pas besoin de savants." Without repudiating capitalism, we had legislatively affirmed that our capitalism had no need of capitalists. In regarding the capitalist simply as a consumer—a favored consumer, to be sure—we had deprived him of his particularized functions.

One may recognize that the functions ascribed to capitalists were and will be serviceable functions, but argue that those functions can be performed as well or even better by some other economic entity than the capitalist—that capitalism can function without capitalists provided it finds a substitute. We shall only stipulate that we apply to the alternative the same test that we applied to the capitalist—namely, that it be capable of reconciliation to social aspirations. For example, corporations are now able to exercise the entrepreneurial function in instances even more satisfactorily than individuals. The laboratories of our great corporations are continuously prosecuting research in chemistry, physics, and mechanics on a scale which no single entrepreneur or entrepreneurial group could emulate. In the important economic function of marketing, magnitude itself seems to make the vast corporation preeminent. These corporations have access to capital, but are not dependent on capitalists. But pursued to its logic, this involves a kind of *numerus clausus* of enterprises. Those now in being would be permitted to expand indefinitely, except as this initial number might be annually reduced by the averages of corporate mortality, until a few hundred corporations would come to represent the state of the arts. That this was the drift seemed to be the conclusion of the Senatorial Committee on the Concentration of Economic Power, whose hearings were held just before the war; and this drift—if such it was—seems to have been accelerated by the processes of the war economy. The same tendency had been noted in other countries in the period between the wars; and there it was generally accepted. Indeed, its logical next step—the consolidation of these corporations into monopolistic cartels—was either tolerated or welcomed except in the United States. We admire big business and we welcome what bigness can offer in the way of material blessings—up to a point; and at that point we have enacted anti-trust laws, prosecuted cartels and suspected monopolies, and have become alarmed over the concentration of economic power. This does not seem to me illogical, or inconsistent; it does seem to suggest that we might find that this alternative to the capitalist likewise conflicts with our social aspirations.

Indeed, the whole of our history indicates an aversion to concentrations of economic power in the hands of a few, whether those few are called nobles, colonial proprietors, capitalists, economic royalists, managers, administrators, or what not. It requires no little ignorance to write an eco-

nomic history of America in terms of the class struggle; it is no effort at all to write it in terms of a horizontal American antipathy to those persons and those institutions which at various times have represented concentrations of economic power. Nevertheless, there is a good deal of evidence that the drift of economic history abroad in the years before 1939 was in the direction of the corporative state—capitalism without capitalists—and that in this country, with the capitalist incapacitated by taxation, there was not a little pressure to repeal the anti-trust acts and permit our economy to follow the European path.

There remains, of course, another alternative. The state might assume the functions of the capitalist. The capacity of the state to furnish the capital formerly provided by the entrepreneurial capitalist is not in question. The capacity of the state to provide the personal quality of entrepreneurship likewise seems to me beyond question, but in the opposite sense. The abstraction known as the state becomes incarnate in the person of the official, an official of the appointive, elective, or civil service type. We will hardly suppose that under state capitalism the office, let us say, of county or municipal entrepreneur will be put on the ballot. The civil service or career type—an admirable character for certain functions of the state—is not prepossessing in the prospective role of entrepreneur. And as for the possible appointive entrepreneur, I doubt very much that one could find any expression of social aspirations that would regard the function of the entrepreneur as suited for inclusion in the patronage lists. The administrative capacity of the incarnate state is, I believe, higher than it is often rated; its entrepreneurial capacity is by no means equal to its administrative capacity. Good intentions are no substitute for professional competence; nor does calling a man a public official automatically endow him with a civic conscience.

It is quite possible that the purely sumptuary aspects of economic equality could be more acceptably handled either under corporative capitalism or under state capitalism, but the concentration of economic power, which hitherto has seemed to be in conflict with our social aspirations, would be complete. You recall the British historian who devoted his life to the compilation of a history of liberty. The history was never written, but Lord Acton will long be remembered for his aphorism, "All power corrupts; and absolute power corrupts absolutely." If he had written all the volumes he planned, he would still be known for that sentence. It is equally applicable to any concentration of economic power, whether in the person of the capitalist, the corporate executive, or the public official.

This analysis suggests that the term "economic equality" embraces two different ideas. One form of economic equalization concerns the narrowing of extreme disparities in individual claims on consumption. This kind of economic equality might be more readily attainable either under the corporate state or under state capitalism than under private capitalism. The other form of economic equality concerns the narrowing of extreme disparities in the individual exercise of economic power. This is certainly more readily attainable under private capitalism than under the concentrations of economic power implied in the corporative state or under state capitalism. I regard these two concepts of economic equality as recognizably distinct if not in mutual conflict. I do not believe both are simultaneously approachable; but I affirm—and this may be no better than a subjective value judgment—that our social aspirations have hitherto rated equalization of economic power as the more important.

The industrial revolution brought capitalism, and until the industrial revolution itself is cancelled, I see no prospect of any other form of economic organization. The world to date has known three types of capitalism—private capitalism, exemplified by the United States and England; the corporative state rather abortively exemplified by Italy and Germany; and state capitalism, as exemplified by Russia.

We are concerned with capitalism as it has evolved in the United States, and as it may evolve here in the future. I have stated my belief and my reasons for believing that in the scheme of American capitalism, the capitalist performed three recognizable and distinctive functions—the assumption of the risk of entrepreneurship; the function of representing proprietary responsibility in corporate organization; and the function of restraining or preventing the development of monopoly.

Our society has expressed its attitude toward the capitalist by taxation that has to a considerable degree brought a more equal distribution of income, but has eliminated or presently will eliminate the capitalist as a focus of economic power. In achieving this latter end, it has not eliminated the concentration of economic power;

it has merely shifted this concentration from a known responsibility to the hands of a small and prospectively smaller number of less identifiable persons. At the same time, in the pursuit of the more equitable distribution of income, it has deprived the capitalist of his capacity to perform the distinctive functions enumerated, but it has not delegated those functions to any other entity or institution. As a consequence of the non-fulfillment of those functions, our economy has shown a marked loss of dynamism, evidenced principally by its inability to provide employment for the annual recruitment of the labor force.

That society can find a substitute for some of the functions of the capitalist is certain. That it can find a substitute for all of his functions, I question; that the substitute will perform those functions more effectively and at less cost, I doubt; that the substitute will encounter conflicts with our social aspirations as serious as, if not more serious than, those met by the capitalist, I believe. At the same time, I am not prepared to deny the political momentum of those social impulses that motivated our schedules of personal taxation immediately before the war.

When the war is over, one can predict with a good deal of certainty that two conflicting factors will exist in our economy. We shall be confronted with a need of high tax revenues and with a demand for reduction in the then existing taxes. Every sort of taxpayer will be organized into one of a multiplicity of pressure groups, each of which will assemble formidable arguments that it is peculiarly entitled to favor, or, as its brief will say, justice. This is entirely proper, and it is to be hoped that the plane of the argument will be above the level of shibboleths, catchwords, and semantics.

Let me cite an example. A distinction is often drawn between taxes designed for revenue and taxes designed for what are called social objectives. Many persons accept it as an article of faith that taxes should never be designed for any purpose other than revenue, and can be so designed. In the post-war period, as I said above, there is good reason to suppose that the aggregate need of revenue will be great, and that the aggregate of taxes will be large. In taxation of that prospective magnitude, and indeed, of much smaller magnitude, any given tax will produce social consequences, designed or undesigned. We are not talking in the conditional mood, but in the indicative; and in the presence of limited options. Given our prospective aggregate of post-war taxation, it is not too much to say that it will be one of the decisive factors in determining the shape of American society a generation hence. But unless we recognize this fact candidly, if we proceed on the theory that taxation is solely a matter of revenue, or of "the most feathers with the least squawk," we will still find that social consequences, designed or undesigned, will be waiting for us at the end of no long road. On the other hand, if we recognize that social consequences are implicit in taxation, we can hope that the objectives will be clearly stated, that they are genuine reflections of social aspirations; and that the means employed are relevant to the desired ends.

The matter of taxation is only one of the points at which a reconciliation must be sought between our social aspirations and the state of the arts. The subject of tonight's discussion is only one aspect of taxation. I brought it to a focus on the capitalist because he illustrates the subject at its most critical, controversial, and complex point. I have tried to describe him in his functional character—the entrepreneur of new undertakings, the representative of the proprietary interest in corporate organization, the invader of monopoly —in short, within the framework of our scheme of capitalism, the professional revolutionist. I have suggested that he offended our social aspirations because there was no rationalized justification of his personal economic stature, in terms of economic egalitarianism or of his reputed economic power. Our legislation has, in my opinion, deterred him from performing his function and has provided no successor, substitute, or alternative. In the absence of those functions, our economic system has exhibited certain recognized faults—a failure to absorb the annual recruits to the labor force and a tendency to drift toward concentration of economic power. The end of the war will find us with the unresolved dilemma of reinstating the capitalist in his functions, under some form more acceptable to our social aspirations, or of providing a substitute which will perform those functions within the frame of social acceptability.

This is not an easy problem, nor in my opinion, is there any quick, facile, and final answer. While I spoke of the reconciliation of social aspirations and the state of the arts as a problem in political economy, political economy becomes opportunism unless it is guided by a social philosophy.

I have devoted this paper to an argument that

our society needs capitalists. I shall conclude it with an argument that our society is in still more urgent need of social philosophers. Throughout this paper American experience has been referred to principally in materialistic terms of the industrial revolution and of capitalism. But America is also the expression of a political philosophy. If ever the Platonic ideal of the philosopher ruler was realized, it was at the founding of America. Never, in one place and at one time, was a nation blessed by such a group of political philosophers as the godfathers of the Republic—Jefferson, Madison, John Adams, Hamilton, Franklin, Rush—farmers, lawyers, printers, surgeons by profession, but all political philosophers, whose wisdom was compounded of experience in the working world and the meditations of the study. America has been less favored with social philosophers—or economic philosophers, if you prefer the term; but we have had as economic philosophers Henry George and Thorstein Veblen; as legal philosophers Justices Holmes and Brandeis. They wrestled with the problems of the reconciliation of social aspirations with the state of the arts; but where are their successors?

Critical appraisal of social aspirations is a function of the philosopher, whether this philosopher is in daily life a lawyer, a farmer, a business man, or a politician. To any given problem of social reconciliation, then, will be several answers, and the one chosen will be the expression of social preference. As Santayana has observed: "Preferences are matters of morals, and morals are a part of politics. It is for the statesman or the humanist to compare the functions of various classes in the state and the importance or timeliness of the various arts."

TAXES AND NATIONAL INCOME

SIMON KUZNETS

University of Pennsylvania

(*Read February 18, 1944, in the Symposium on Taxation and the Social Structure*)

TAXES AND OUTLAYS

THE individual and the business firm tend to view taxes as drafts upon their stock of economic claims, as a surrender of part of the power which they could otherwise exercise in the markets of economy. Since, with few exceptions, payment of a tax is not connected with specific service to the payor and is backed by legal sanctions rather than by economic attractions, the individual and the firm do not usually view taxes as charges for goods purchased. There is thus lack of association in the citizen's mind between the hand that pays taxes and the hand that receives benefits from the government. Nor is there clear realization that taxes are only one of the ways by which the government can withdraw from individuals and firms part of their economic power; that it can do so, for example, by borrowing, by printing money, or by so affecting the production system as to diminish the supply of goods available for free purchases.

This tendency to view taxes as a necessary evil which, like death, one must bow to but avoid as long as possible; to disregard their role as a prelude to some useful government activity; and to overlook the fact that they are only one of several methods by which diversion of resources to government use can be accomplished, is of great practical importance in policy decisions. It sets limits to the kind and amount of taxes that can be collected at any given time; and it permits, altogether too easily, recourse to other methods of government financing that are less unpleasant to the individual payors but perhaps more dangerous to the welfare of the country. And it has its advantages in stimulating critical scrutiny of the uses that the government proposes to make of resources withdrawn from direct control of individuals and business firms.

But in considering the relation of taxes to national income, as part of a general discussion of their bearing on social structure, we must perforce take a broader view than the one just attributed to the taxpayer. We must keep in mind *all* the methods by which government can reduce consumption or investment by individuals and firms; and recognize that the way the government spends money it collects can raise national income not only by the direct addition of the government's own production of commodities and services, but also by indirect influence upon consumption and investment propensities of individuals and firms. The problem is thus one of trying to ascertain the *net* balance of the depressive effects of taxes and other government collections and the expansive effects of their disposal upon the magnitude, structure, and trends of national income.

In addition to taxes, government may also receive revenues that are a charge for specific services rendered, for example, the 3-cent stamp for mailing a letter. Such non-tax revenues are of relatively minor importance in this country, but loom large in others where the government owns and manages a good proportion of the country's industries. In addition to total revenue, the sum of tax and other collections, government can of course borrow money; and it does so on a substantial scale even in peacetime, let alone during the years of a major war. We may designate the sum of revenues and borrowing the total monetary intake of government.[1]

With the proceeds of this total monetary intake the government incurs expenditures that are essentially of two types—the classification being economic in character. The first, commonly designated transfer expenditures, comprises cases in which the government pays the money collected or borrowed from one group of the population, to another group within the country or

[1] We should note the exclusion here of other types of government collections which have relatively little use in this country. The government can, of course, print money; and it can extract resources from members of the nation by confiscation, sometimes on the pretext that one or another group in the body national is not entitled to the possession of its economic goods. The reason for excluding such policies from consideration here is that they are not in consonance with the basic assumptions of the society we are discussing; although some types of borrowing are not much different in their economic effects from printing paper money.

outside, *without at the same time requiring these recipients to deliver to the government either their services or their products.* Typical illustrations are the payment by the government of interest on government debt, and of pensions to veterans. The second type, which we may designate purchase payments, includes outlays by governments either for services of people, to work directly under government auspices, or for products of private firms to be used by the government in its operation. Typical illustrations are payments of salaries to government employees or for materials to be used by government offices, the Army, the Navy, etc. Finally, in addition to these two types of expenditures, governments can use the money to repay debt or to set up credit balances. These are not expenditures in the ordinary meaning of the term; but from the economic standpoint they are quite similar to transfer expenditures as defined above.

The distinctions frequently made between transfer expenditures and purchase outlays are that: (a) a transfer payment is not evidence of any real production corresponding to the payment, and the transaction is mere shuffling of economic claims; (b) a transfer payment means no withdrawal of real resources from use by the private sector of the economy. Evidence either of real production or of withdrawal of real resources is supposed to be true of purchase payments. A close analysis, however, reveals that neither distinction is fully valid. Some types of transfer payment may be evidence of real production, that is, of yield of real factors. Thus, payment of interest on public debt, like payment on any investment of capital in the past, is evidence, at any rate for the purpose of income measurement, of the current yield of such capital (the balancing item being the net profit or loss of enterprise using that capital). Hence, payment of interest on public debt, while a transfer expenditure as ordinarily classified, is, in income measurement in this country, taken as evidence of real production occurring in the economy.[2] Similarly, it may be disputed that transfer expenditures cannot involve withdrawal of real resources. A payment by the government to a group of pensioners or debt-holders may put them in an economic position in which they are not forced, as they would otherwise be, to offer their services on the labor market. In that sense, transfer expenditures may mean withdrawal of *potential* real resources from possible control by private enterprise. The true kernel of the distinction lies in the fact that transfer outlays do not mean the placing of *existing* real resources within the control of the government for direct use under government auspices; whereas purchase outlays necessarily mean that the government absorbs products of business enterprises and services of people.

The purchase outlays by government begin to approximate, but are still different from, the value of government's contribution to the national product. What is the net addition to goods at the disposal of the nation which the governments make *directly?* The outlays by government on commodities and services are obviously too large for the purpose of measuring the *net* contribution of government, since they include *products of other industries*, the yield of productive factors engaged in these other industries on their specific tasks. The true contribution of government or its net value product is either the value of services rendered by productive factors engaged in the government industry proper, or such yield of productive factors in the economy as cannot be traced to any specific industry but can best be allocated to past investment by the government for the benefit of the country at large. Thus, the contribution of government to net national product is defined as the sum of: (a) the compensation of the productive factors directly engaged in work for the government, largely payments to its employees; (b) net interest on government debt; (c) by analogy with private enterprise, the net profits or loss of the government viewed as an industry, that is, its net savings measured by a comparison of the change in its debt with the change in its real assets.[3]

[2] In other countries (*e. g.*, England) past practice has been to consider some part of public debt as "deadweight" (particularly debt incurred on account of war outlays), and to treat interest on it as "pure" transfer.

[3] Where such measurement of government savings is undertaken, all payments by governments to individuals are included, so that in addition to (a) and (b) the estimates include also pensions, relief payments, etc. Otherwise, the net value product of government would be underestimated, under (c), by the possible increase in debt associated with such payments.

The effect of this procedure is to value current activities of government (total ex. additions to real assets) at the amount of taxes paid for them. This has been the author's practice, and represents an attempt to resolve the vexing problem of valuation of government activity on a basis that is most comparable with the market valuation basis of the private sector of the economy. Other procedures, em-

COMPARISONS WITH NATIONAL INCOME

We have distinguished, in the flow of means of payment to the government, taxes, other revenues, and borrowing; in the outflow from government, repayment of debt or addition to balances, transfer payments, and purchase outlays; and we have defined the net value product of government activity. The totals that can thus be derived—taxes, total revenue, total monetary intake, total monetary disbursement, total expenditures, transfer outlays, purchase outlays, net value product—may each be conceived as part of a larger country-wide total; as the numerator for which an appropriate denominator should be secured, in order to permit us to measure the relative weight of government, either as collector and disburser of funds, or as claimant upon real resources and producer of real goods.

To simplify the discussion, we shall reduce the number of numerators for which to seek appropriate denominators in some variant of national income. We propose to deal below with only four totals expressive of government transactions: (a) taxes; (b) total money intake, modified to include *net* rather than gross borrowing; (c) purchase outlays; (d) net value product of government.[4] What are the appropriate measures of country-wide total activity with which these four summaries of government activity can properly be compared? In answering this question, we shall have to specify in each case the basic problem upon which the comparison is assumed to shed light; and to select a denominator of which the numerator can logically be considered a constituent part.

(a) The total of taxes is of particular interest as a measure of the draft which the government makes upon incomes of individuals and business firms, without providing each with a specific quid pro quo. In so far as these incomes of individuals and business firms are received by them in return for useful economic activity, and in so far as they provide incentives for whatever economic contribution the individuals and enterprises make, reduction of such incomes by taxation may involve a reduction in stimuli. In this light the proper total with which the amount of taxes collected by the government should be compared is obviously the sum of net incomes of individuals and private enterprises. Net incomes of individuals in this connection may be defined as their total income, minus specific occupational expenses incurred in the pursuit of economic activity. Net income of enterprises may be defined as net revenue, after covering all the fixed and current costs but before payment of income and profit taxes and of dividends.[5] The aggregate, which is statistically quite close to national income as ordinarily measured, may be designated the net private income aggregate.

(b) The total monetary intake of government comprises, in addition to taxes, other revenues which on a net basis are, at least in this country, a minor proportion of total government intake at any time; and net borrowing, which at times forms a large share of the flow of means of payment to the government. It must be noted that in the long run, in terms of decades, net borrowing is only a limited share of the total flow of money to the government.[6] Hence, a distinction between taxes and the total monetary intake as here defined is significant only for the shorter run; primarily for years in which extraordinary needs of the government make it impossible to finance a large proportion of government outlay from any source except through borrowing from individuals and enterprises.

What is the total at the disposal of individuals and enterprises upon which total monetary intake of government represents a draft? At any given time, the sources from which government can finance whatever outlay it wishes to incur,

ploying a *cost* basis of valuation of government activity and different assumptions concerning the distinction between government services to individuals and to business firms, are in wide use. On this subject see extensive discussion in the several volumes of *Studies in Income and Wealth* (Nat. Bur. Econ. Research) **1**, 1938; **3**, 1940; **6**, 1943.

[4] We may note briefly the reasons for omitting from further discussion the other totals just distinguished. Total revenues in this country are quantitatively not greatly different from taxes, since other revenues form only a minor proportion of all government receipts. Total expenditures are equal to total monetary intake of government, when only net borrowing is included, *i. e.*, only the difference between gross borrowing and additions to balances plus reduction of debt. The omission of both total monetary intake and total monetary disbursement is largely for the purpose of disregarding the "wash" transactions, even though some effects of these are suggested in the discussion below.

[5] The implication is that the income and profit taxes of corporations are not transferred to individuals—an assumption widely accepted in taxation theory but subject to some exceptions. To the extent that such taxes are so transferred, they are already covered under individuals' incomes and should not be included once again in the net private income aggregate.

[6] See the appropriate ratios in table 5 of National Income and Taxable Capacity, *Am. Econ. Rev.*, suppl.: 72, March, 1942.

include (i) net incomes of individuals as defined above; (ii) net incomes of enterprises, again as described above; and, in addition, (iii) such returns of current activity as in the short run do not have to be re-invested to replace capital consumed. This third category comprises the part of final product value that covers depreciation and depletion of durable goods sustained in the productive process. If necessary, re-investment can be delayed and the funds made available to the government, without any necessarily deleterious effects upon the productive performance of the economy for the short run. We may designate the corresponding total the gross private income aggregate, gross with respect to depreciation and depletion of fixed assets.[7]

(c) The particular interest attaching to the total of purchase outlays lies in the fact that it measures the value of commodities and services consumed by the government in the process of its activity. As distinct from the two totals already discussed, which are on the level of monetary flows, purchase outlays emphasize treatment in terms of real goods and resources.

Two approaches may be taken to this total. In the first, we may wish to know the comparative magnitude of government activity as measured by its real cost; in the same way as we may wish to measure the magnitude of construction activity by the amount of materials and labor that it consumes. The wider total with reference to which the relative weight of government activity should then be measured is a total that would add, for the various industries distinguished, the respective values of commodities and resources consumed. The difficulty is that the size of any such total would depend upon the character of the industrial classification used: the larger the number of industrial divisions, the larger the total because the more extensive the duplication caused by repeated counting of products of one industry consumed by another. In this sense, there is no uniquely determined country-wide total, in comparison with which the relative weight of government activity can be established.

This difficulty may be overcome by a second approach, which compares the real cost of government activity, all classified as "finished" production, with the total of finished products turned out by the economy. To make the denominator comparable with the numerator, as a source upon which the latter may be viewed as a draft, we must treat all government production (measured at purchase outlays) as finished output; and thus include it fully with the two other major components of finished output—consumers' goods flowing to households and individuals, and private gross capital formation. The total is then national product gross of: (a) consumption of private durable capital; (b) consumption of government products, *i. e.*, that part of purchase outlays that is covered in the value of consumers' outlay and private gross capital formation. It is with this total that we may compare governmental expenditures on goods and services.[8]

[7] There is common misunderstanding to the effect that government can borrow not only out of current savings, but also out of past savings in the sense of using proceeds of conversion of privately held existing assets into government bonds. Actually, of course, previous investments by individuals or enterprises can be transferred to government only if the latter purchases already existing assets from private parties. Another channel of possible utilization of past savings, activation of balances otherwise held idle, may also be mentioned, although it again is limited. The possibilities of both methods of financing of government out of past savings are exceedingly narrow, so that no major error is involved in the statement that the sole source of government borrowing is current savings, gross or net. See, however, second paragraph of footnote 8.

[8] This obviously is a peculiar national product total, taken gross in a specific way corresponding to the area for which the volume of activity is taken gross (*i. e.*, measured by expenditures on commodities and services, rather than by net value product). Were we to deal with the purchase outlays of the steel industry, rather than of government, the denominator would be national product gross of: (a) consumption of durable capital outside the steel industry; (b) consumption of steel industry's product, *i. e.*, that part of the industry's purchase outlay that is covered in the value of consumers' outlay or gross capital formation outside the steel industry. In other words, instead of making national product gross with reference to *all* industries in the industrial classification, we do so only with reference to the particular industry whose purchase outlays are studied in the numerator. The significance of such a denominator and of the resulting ratio clearly depends upon the propriety and meaning of emphasizing the gross volume of activity in a given industry, as measured by its total expenditures on commodities and services; and accordingly of treating this gross value as "finished" production.

Another point to be noted is that if purchase outlays include large drafts upon already *existing* assets, the denominator cannot be confined to *current product* even if including the consumed part of durable capital. In this case, it may well be better to derive two ratios—one for draft upon current products, the other for draft upon capital (see suggestion by Dr. Raymond Goldsmith in *Studies in Income and Wealth* (Nat. Bur. Econ. Research) **6**: 62–67, 1943. This problem, however, does not arise in this country where little draft is made by government purchase outlays upon capital stock (beyond the currently consumed part of durable capital).

(d) The net value product of government is its contribution to the net national product, commonly designated national income. The proper definition of this denominator, in terms of final goods, is the value of all commodities and direct services produced, minus the value of *commodities* consumed in the process of production. This national product total differs from the total just suggested under (c), by its being net of the consumption of durable capital and of the consumption of the products of government activity. At the payments level, national income is equal to the sum of all payments by enterprises to individuals, undistributed net profits of business enterprises after all taxes, and net savings of governments. This national income total differs from the net private income aggregate defined above in two respects: (i) it excludes income taxes paid by business firms; (ii) it includes net savings of government.

With the four numerators and the corresponding denominators established, and the resulting four ratios described, we are in a position to summarize the available statistical evidence concerning the magnitude of these ratios, their longer term trends, and their levels and changes during recent years. A few preliminary comments are in order to place the evidence within the proper framework.

First, the over-all ratios described are in and of themselves summaries that may conceal aspects of the comparison more important than the ratios themselves. For example, the proportion of taxes in the net private income aggregate may be equal in two countries, or in the same country in two different years, and yet the structure of the tax system in its impact upon different classes of individual recipients or different classes of enterprises may be vastly different. Indeed, one of the most intricate and fruitful groups of studies in public finance is that of different types of taxes, designed to minimize their restrictive impact upon the economy and to permit at the same time the draft upon current incomes that is necessary to sustain indispensable government activities. Similarly, the ratios of total monetary intake to gross private income aggregates, or of purchase outlays and net value product to the corresponding national product totals, fail necessarily to reveal possible differences that may be decisive in any thorough causal analysis.

Second, there is a significant difference between comparisons that treat both government and over-all activities on the level of monetary flows (such as those under (a) and (b)) and those that treat them in terms of real goods, whether they be productive resources or finished products (such as in (c) and (d)). In those that deal with monetary flows, it is perfectly legitimate to compare dollar totals without worrying about the fact that the prices that governments on the one hand, and private enterprises and individuals on the other, pay for real goods, may differ markedly. In comparisons of real goods, the differences in the structure of markets and pricing schemes between government and private purchases are both significant and exceedingly difficult to correct for.

Finally, the definitions given above of national product and of the income aggregates slide over many thorny problems in distinguishing and measuring certain components. The difficulties of defining, let alone measuring, such components as "net savings of business enterprises" or "net savings of governments" have been the subject of long and none-too-conclusive technical discussions. They cannot be treated here, even though they must be briefly mentioned in connection with some of the statistical comparisons that follow.

PRE-WAR TRENDS

Table 1 summarizes movements in the ratios of taxes and outlays to income and product aggregates for as long a period before the present war as can be roughly measured with available data. Granted the approximate character of both the numerators and the denominators, the broad indications of relative levels and of trends are clear and beyond doubt. They show: first, that prior to World War I the relative magnitude of government activity, measured in the several totals distinguished, was quite minor, well below 10 percent; second, that after World War I there was an appreciable rise in the percentage which the various totals expressive of government activity formed in the income and product aggregates, but even so the shares were moderate, running slightly above 10 percent for taxes and expenditures, and well below 10 percent for purchase outlays and net value product; third, that the depression that followed 1929 was accompanied by an increase in the ratios of taxes, expenditures, and net value product to the country-wide income or product aggregates; fourth, that the recovery from the depression, in so far as such recovery took place before the

TABLE 1

LONG-TERM PRE-WAR TRENDS IN RATIOS OF TAXES, GOVERNMENT OUTLAYS, AND GOVERNMENT NET VALUE PRODUCT TO CORRESPONDING NATIONAL INCOME DENOMINATORS

Years	Percentage of Tax Collections to National Income	Percentage of Total Expenditures to		Percentage of Purchase Outlays to		Percentage of Net Value Product to National Income
		National Income	Gross Income Aggregate	National Income	Corresponding Gross Product	
	(1)	(2)	(3)	(4)	(5)	(6)
1860	4.3	N. D.	N. D.	N. D.	N. D.	[5]2.3
1870	10.0	N. D.	N. D.	N. D.	N. D.	[5]4.0
1880	8.7	N. D.	N. D.	N. D.	N. D.	[5]3.9
1890	7.3	6.0	5.4	N. D.	N. D.	[5]4.5
1902	5.8	[1]6.0	[1]5.4	[1]4.2	[1]3.4	5.1
1913	6.5	7.6	6.8	5.7	4.6	5.2
1922	12.3	[2]12.4	[2]11.2	N. D.	N. D.	8.5
1928	11.6	13.5	12.1	[3]9.4	[3]7.5	7.9
1934	17.1	26.5	23.8	N. D.	N. D.	12.3
1938	22.7	25.8	23.2	[4]14.9	[4]11.9	[6]11.5

[1] Comparison for 1903.
[2] Comparison for 1923.
[3] Based upon ratio for 1929.
[4] Based upon ratio for 1937.
[5] Comparison for preceding years, viz. 1859, 1869, 1879, 1889.
[6] Comparison for 1937.

Sources:

Col. 1 and 2: From the author's article in *American Economic Review*, Supplement, National Income and Taxable Capacity, table 1, p. 63, March, 1942.
Col. 3: Col. 2 reduced by one-tenth.
Col. 4: Derived by applying to column 2 percentages of "other" to total expenditures, *ibid.*, table 4A, p. 68.
Col. 5: Col. 4 reduced by two-tenths.
Col. 6: From Robert F. Martin, *National Income in the United States, 1799–1938* (New York, 1938), table 40, p. 87. Represents ratios of production income from government to total realized national income.

outbreak of World War II, did not result in an appreciable lowering of the ratios of taxes, or expenditures, or net value product, so that in a sense we entered this war with the government activities playing a part in the country-wide income or product totals close to their secular peak levels.

Against these broad tendencies common to the ratios presented in table 1, differences in the respective trends are quite minor; the main differences are in the levels. The percent of tax collections (the total comprising all taxes, direct and indirect, customs, excises, etc.) moves from the very low pre - Civil War levels of 4 percent to 23 percent in 1938. The ratio is taken to national income (exclusive of government savings) rather than to the net private income aggregate; but the difference in the denominators is quite minor, would not affect the trend of the ratio, and would modify the level itself by only one or two percent in recent years. The rise in the ratio is due to a fairly steady increase in the relative shares of state and local taxes, accelerated after World War I, and to a less consistent rise in Federal taxes, in which the war periods give rise to prominent cycles, the peak associated with the war and the ratio declining somewhat after the war but never returning to pre-war levels. The only significant addition to this picture is that the depression following 1929 had the same effect on the Federal tax ratios as the wars, causing a rise in the ratios similar to those noted after 1860 and after 1915.

The second ratio in table 1, that of total expenditures to national income, is quite close to that of taxes in the years before the recent two decades (column 2). This ratio should properly be taken to the gross private income aggregate; and a rough reduction by one-tenth from column 2 to column 3 takes account of the order of difference in the denominators. The trend, in so far as the available data permit its establishment, is quite parallel to that of the ratio of taxes to national income.

Purchase outlays should be related to national product gross of consumption of durable capital and of government products, a total that tends to run about 20 percent larger than national in-

come as ordinarily measured; hence the crude adjustment from column 4 to column 5. The level of the ratio of outlay expenditures is naturally lower than that of the ratio of total expenditures, because the former account for roughly two-thirds of the latter, the other third being accounted for by transfer outlays. Over the three and a half decades preceding the present war, the ratio in column 5 almost quadruples, a rise only somewhat smaller than that of the ratio for total expenditures in column 3.

On the percentage of net value product to national income, we have data extending all the way back to the beginning of the nineteenth century. The available estimates of national income exclude government savings. But the item is relatively minor barring years of exceptional prosperity or depression, or of extraordinary outlays in connection with major wars; and the effect of its omission on the long-term comparison is thus likely to be negligible. The available series shows that the rising trend in the ratio of government net value product to national income is of even longer standing than is indicated by table 1, which covers a period of eight decades. The trend is almost continuous from ratios which are at 1 percent in the beginning of the nineteenth century, to roughly 2½ percent by 1860, to 5 percent by the beginning of the twentieth century, and to 12 percent in the years immediately preceding the present war.

The particular point that calls for comment is the large difference in the level of this percentage, which measures the relative contribution of government activity to net national product, and the percentage that is roughly twice as large that measures the draft of tax collections upon the net private income aggregate. The reason, of course, is that tax collections are designed to finance not only payments by governments to the productive factors under its own auspices, but also transfer expenditures of the type that means shuffling of economic claims among various groups of income recipients, and payments for products of other industries consumed in carrying on government activity. For any other industry there would also be a similar difference in the ratios to national income between the total *charge* of the industry for its product, *i. e.*, total sales value of its finished articles and the *net value product*, *i. e.*, compensation of the productive factors engaged in that industry. The peculiar characteristic of government, as compared with other industries, is that more than others it incurs transfer expenditures, *i. e.*, reduces the purchasing power of some groups of the population for the benefit of others, without a corresponding market exchange of economic goods. In any consideration of the effects of the secular rise of government's share in total economic activity, either past or prospective, this double function of government, both direct production of goods and rearrangement of the income distribution, must be clearly kept in mind.

RATIOS FOR THE WAR YEARS

It may be asked whether the trends observed in the decades up to the present war towards an increasing proportion of taxes, outlays, and government net value product, are likely to continue in the post-war future. No certain answer can be given to this question; but reasonable hypotheses can be advanced. Before discussing them, however, it is of interest to consider the ratios of taxes, outlays, and government net value product to the relevant over-all denominators during the recent war years.

Table 2 presents a comparison between taxes and total outlays and the corresponding income aggregates, annually from 1939 through the first half of 1943. For these years, particularly 1942 and 1943, it becomes important to measure more precisely the proper denominators, *i. e.*, the net and gross private income aggregates, since they differ materially from national income as ordinarily measured. Thus, for 1942 the net private income aggregate is larger than national income, as customarily measured, by 7 billion dollars or 6 percent. The gross private income aggregate, which includes in addition depreciation and depletion and net business tax accrual, is, in 1942, larger than national income by 20 billion dollars or well over 15 percent.

Two conclusions stand out clearly in table 2. First, the ratio of taxes to the net private income aggregate shows no great rise, the percentage increasing moderately from 18 in 1939 to 20 in 1942 and 23 in the first half of 1943. Thus the draft represented by total taxes upon the income receipts of individuals and net undistributed profits of corporations (before tax) has kept pace with or only moderately exceeded the rise in the volume of these income receipts. This is contrary to the popular impression that the magnitude of taxes, relative to the incomes of indi-

TABLE 2

PRIVATE INCOME AGGREGATES, TAXES, AND OUTLAYS, 1939–1942
(Billions of Dollars)

	1939	1940	1941	1942	1st Half 1943[1]
1. Income payments to individuals	70.8	76.5	92.2	115.5	137.6
2. Contributions to social insurance[2]	2.0	2.1	2.6	3.3	3.8
3. Corporate savings, adjusted	1.5	3.2	2.6	4.0	6.0
4. Adjustments for statistical discrepancies	—	−0.9	−0.8	−0.3	+3.8
5. Business income and profits tax	1.1	1.5	2.8	7.3	8.0
6. Net private income aggregate (1+2+3+4+5)	75.4	82.4	99.4	129.8	159.2
7. Depreciation and depletion	6.2	6.4	6.9	7.8	8.0
8. Net business tax accrual	0.3	1.3	4.6	5.3	5.8
9. Gross private income aggregate (6+7+8)	81.9	90.1	110.9	142.9	173.0
10. Tax collections (budget receipts)	13.6	14.9	18.3	26.1	36.4
11. Total outlay (budget expenditures)	17.8	18.3	27.6	64.0	96.0
12. Percentage of taxes to net private income aggregate (lines 10 and 6)	18.0	18.1	18.4	20.1	22.9
13. Percentage of outlay to gross private income aggregate (lines 11 and 9)	21.7	20.3	24.9	44.8	55.5

[1] Seasonally adjusted annual rate.
[2] Included fully, although part is contributed by employers; and this part, while entering the gross private income aggregate, should not be included in the net private income aggregate.

Sources:

Line 1: *Survey of Current Business*, March, 1943, table 12, p. 21, line 5, and *Survey*, August, 1943, table 3, line 1.
Line 2: *Survey*, August, 1943, table 7, p. 13, line 4.
Line 3: Corporate savings from *Survey*, August, 1943, table 6, p. 13, line 11. Adjustments for additions to business reserves, capital outlay charged to current expense, and inventory revaluation from *Survey*, August, 1943, table 4, p. 12, lines 4, 5, 6.
Line 4: *Survey*, August, 1943, table 4, p. 12, line 7.
Line 5: For 1939–1942, *Survey*, March, 1943, table B, p. 25—sum of Federal corporate income tax receipts (line 2) and state and local corporation income taxes (line 31). For 1943, the 1942 total multiplied by the ratio of business tax and non-tax liabilities in first half of 1943 (doubled to annual total) to annual total of same item in 1942 (see *Survey*, August, 1943, table 4, p. 12, line 2).
Line 7: *Survey*, August, 1943, table 4, p. 12, line 3.
Line 8: *Survey*, March, 1943, table 6, p. 19, line 3. For 1943 assumed at the same ratio to line 5 as in 1942.
Line 10: For 1939–1942, sum of Federal net budget receipts (*Survey*, March, 1943, table B, p. 25, line 13) and total state and local budget receipts (*ibid.*, line 36). For 1943, based on ratio of 1943 first half total of (business tax plus non-tax liabilities minus net business tax accruals plus individual taxes) to similar total in 1942, the ratio applied to entry for 1942 in line 10. For entries underlying the ratio see *Survey*, August, 1943, tables 4 and 7, and line 8 of present table.
Line 11: For 1939–1942, sum of Federal expenditures (*Survey*, March, 1943, table A, p. 25, line 10) and state and local budget expenditures (*ibid.*, line 23). For 1943, ratio of government expenditures on goods and services in first half of 1943 (doubled to cover year) to total in 1942 (*Survey*, August, 1943, table 5, p. 13, line 2) applied to entry for 1942 in line 11.

viduals and net earnings of corporations, has increased strikingly during the war years.[9]

By contrast, the proportion of total outlay to the gross private income aggregate rose markedly, from 22 percent in 1939 to 55½ percent in the first half of 1943. This ratio measures the relative draft of both government taxes and borrowing upon incomes of individuals and corporations, inclusive of their depreciation and depletion and other reserve accumulations during the year. In combination with the first conclusion, this rise in the outlay-to-gross-income-aggregate ratio clearly shows that the financial burden of increased government needs was carried not by any appreciable increase of the *relative* tax burden; but rather by an increase in the relative draft by

[9] This conclusion is reenforced if we distinguish between taxes falling on individuals' incomes and the corporate income and excess profit taxes. Omitting the latter, the tax bill rises from 12.5 billion dollars in 1939 to 28.4 billion dollars (annual rate) in the first half of 1943. (See lines 10 and 5 of table 2.) The ratio to income receipts of individuals (lines 1 and 2) is at 17.2 percent in 1939 and at 20.1 percent in the first half of 1943—a rise in the ratio smaller than that shown for all taxes in line 12 of table 2.

the government, via borrowing, upon the savings of individuals and firms.

The attempt to calculate the same ratios, not for money flows such as are represented by taxes, outlays, and income payments, but for real resources or goods (even if valued in terms of money), is difficult in a war year, because of two basic problems: (a) that of measuring real net product of government activity, particularly government net savings; (b) that of correcting for the differences in level between prices paid by government and those paid by individuals and business enterprises. The first problem requires an assumption as to what part of war production can be treated as additions to the government inventory of real assets, to be used as an offset against the increase in government debt. The second is really a question as to the relative efficiency of resources in government production, compared with those in private production.

The upper part of table 3 presents a comparison of purchase outlays with the corresponding variant of gross national product, *i. e.*, one gross of consumption of durable capital and of

TABLE 3

NATIONAL PRODUCT, PURCHASE OUTLAYS, AND GOVERNMENT NET VALUE PRODUCT, 1939–1942

	1939	1940	1941	1942	1st Half 1943[1]
1. Consumers' outlay	61.7	65.7	74.6	82.0	89.7
2. Private gross capital formation	10.9	14.7	19.0	8.0	−0.7
3. Government purchases of goods and services	16.0	16.7	25.7	61.7	92.3
4. National product, gross of consumption of durable capital and of government product (1+2+3)	88.6	97.1	119.3	151.7	181.3
5. Ratio of purchase outlays to national product gross of consumption of durable capital and of government products (lines 3 and 4)	18.1	17.2	21.5	40.7	50.9
6. Direct government services to individuals	3.1	3.0	2.8	2.7	2.4
7. Flow of goods to consumers (1+6)	64.8	68.7	77.4	84.7	92.1
8. Non-war capital formation, private, net	4.7	8.3	12.1	0.2	−8.7
9. Net construction, public, non-war	1.3	0.9	1.0	0.3	−0.4
10. Net non-war capital formation (8+9)	6.0	9.2	13.1	0.5	−9.1
11. Additions to inventory of war products	0.9	1.6	7.5	29.5	48.8
12. Net national product (7+10+11)	71.7	79.5	98.0	114.7	131.8
13. Government income payments	10.0	10.3	11.5	18.4	26.8
14. Net additions to government capital (9+11)	2.2	2.5	8.5	29.8	48.4
15. Net budget deficit, all governments	4.3	3.4	9.2	38.0	59.6
16. Net value product of government (13+14−15)	7.9	9.4	10.8	10.2	15.6
17. Ratio, government net value product to net national product (lines 16 and 12)	11.0	11.8	11.0	8.9	11.8

[1] Seasonally adjusted annual rate.

Sources:

Line 1: *Survey*, August, 1943, table 5, p. 13, line 16, and table 1, p. 12, line 14.

Line 2: *Ibid.*, table 5, line 8, and table 1, line 8.

Line 3: *Ibid.*, table 5, line 2, and table 1, line 2.

Line 6: Valued at 1939 direct taxes paid by individuals, modified for later years by ratio of Federal non-war, state and local expenditures in the given year to the corresponding total in 1939. Expenditure data for 1939–1942 in *Survey*, March, 1943, table 10, p. 20, lines 5 and 6; for 1943, in *Survey*, August, 1943, table 1, p. 12, lines 5 and 6.

Line 8: Gross private capital formation (see line 2 above) minus depreciation and depletion (see *Survey*, August, 1943, table 4, p. 12, line 3).

Line 9: See *National Product, War and Prewar*, etc. (Nat. Bur. Econ. Research, pap. 17): table 1, February, 1944.

Line 11: Based on subtracting from gross war production current consumption of war products—see table 3, line 8, in *Treatment of War Production in National Product* (to be submitted to April, 1944, meeting of Conference on Research in Income and Wealth).

Line 13: From *Survey*, March, 1943, and from Department of Commerce (for 1943 first half). Figure for 1942 raised by 2 billion dollars to allow for subsistence of members of armed forces (average of 4 million men multiplied by a crude estimate of $500 per capita). Similar adjustment for 1943 at 3 billion dollars.

Line 15: For 1939–1942—sum of budget deficits and surpluses for Federal, state, and local governments (see *Survey*, March, 1943, table 6, p. 19, lines 12 and 15). For 1943, from lines 10 and 11 of table 2 above.

government products. All the totals are taken directly from the Department of Commerce data, and, to avoid confusion, they follow its concepts.[10] This particular comparison avoids the problem of measuring government savings involved in that of net value product of government with net national product, a problem resolved in the lower part of table 3. Net additions to war capital inventory are calculated by subtracting from total war production current consumption based upon assumed life periods for broad categories; and these together with the value of net public construction (again gross minus depreciation) yield total additions to the real assets of government—to be used as offsets against the increase of public debt in the calculation of government net savings. Both comparisons in table 3 neglect the problem of differences in level between prices paid by government and those paid in the private sector of the economy, by using totals in current, unadjusted prices.

The ratio of purchase outlays to the corresponding variant of gross national product increases from 18 percent in 1939 to 51 percent in the first half of 1943. This rise, concentrated in the years that follow 1941, is clearly associated with war outlay and is quite similar to that of the ratio of total expenditures to the gross private income aggregate in table 2.

By contrast, the ratio of government net value product to net national product shows relative stability over the period, with a slight drop in 1942. The reason is that the basic measure of government net value product in current prices is the volume of taxes. With the taxes increasing only slightly in proportion to net national product, and with an increasing share of taxes going to pay for the current consumption of war goods, which does not enter the net value product of government, the latter would naturally show a stable or declining ratio to total national product. To state the result somewhat differently: as valued by society in terms of taxes that society was willing to pay, the net contribution of government to the total heap of net production increased at a rate barely equal to that of the increase in the total heap itself.

Clearly, this ratio, like the ratio for purchase outlay, is affected considerably by differences in the price level between the numerator and the denominator. It is fair to say that in the comparison of purchase outlays with the corresponding variant of gross national product, the numerator is valued at higher price levels than the denominator, especially in years like 1942 and 1943; so that in terms of comparable price levels, the ratio in 1943 would not be at 51 percent, but lower by perhaps a quarter, bringing it down to roughly 40 percent. In the comparison of government net value product and net national product, as measured here, the implicit price index of the numerator is probably lower than that of the denominator. No rough adjustment can be suggested; but it is not likely to modify the indication of stability of the ratio for 1939–1943, as compared with the rise in the ratio of purchase outlays to the corresponding variant of gross national product.

[10] Although the total it uses for expenditures of government on commodities and services is too large from our viewpoint in its inclusion of interest on public debt. The correction, however, would be relatively minor.

PROSPECTS FOR THE FUTURE

In considering whether the pre-war trends will persist in the post-war future, with a consequent further rise in the shares of taxes, government outlays, and net value product in country-wide income and product aggregates, certain hypotheses can be safely adopted as plausible.

There seems little question that the *absolute* levels of the several totals expressive of government activity will, in the first post-war decade or two, be much higher than they were before the outbreak of the present war. Some of the factors that made for a rise in the absolute volume of government activity in the past will, if retarded during the war years, tend to return when peace comes. The rise in outlays by cities on various types of urban service, in outlays by states in connection with highway systems and education, and in expenditures by the Federal government in connection with social security and public works, is likely to be resumed in the post-war future. In addition, the war itself will leave its usual residue of increased obligatory Federal expenses, in larger interest payments on public debt, disbursements to veterans, and most likely a much larger outlay on peacetime military expenditures than was customary in this country in the pre-war years.

The magnitudes of the post-war government budget cannot be estimated precisely now. But even rough guesses suggest that owing to larger interest payments on debt, larger peacetime military expenditures, larger payments to veterans, and a wider social security system, the Federal budget in the first post-war decade is

likely to be from two and a half to three times as large as the pre-war budget of 8 billion dollars, with a rough level of 25 billion dollars per year. We may also expect some increase in the state and local budget from the 9 billion dollars level, prevailing during the pre-war years, to 10–11 billion dollars. This, in very rough terms, yields a post-war government budget of about 35 billion dollars per year, twice as large as that in the last years before the present war. Under the assumption of no deficit financing, this would also mean a tax bill of equal magnitude, disregarding any possible liquidation of public debt; or a tax bill two and a half times as large as that immediately preceding the war.[11]

Whether the *ratios* for the various totals expressive of government activity, as distinguished from the absolute volumes themselves, will reach in the post-war decades a higher secular level than before the war, depends, of course, upon whether or not the denominators, *i. e.*, the various income and product aggregates, are likely to rise at the rates suggested for total outlays and taxes. The budget and tax comparisons just made between pre-war and post-war do not take account of any price rises. It is unlikely that the income and product aggregates, at constant prices, will reach a level in the post-war decades two or two and a half times as high as the level in 1939. Even the most optimistic assumptions do not suggest an increase between 1939 and the post-war decade of more than 50 percent, *i. e.*, a national income of 120 billion dollars; and a more moderate estimate would probably allow a rise of only 25 percent. This means that the various ratios expressive of the proportion of government activity in national income are likely to rise in the post-war decades to levels appreciably higher than those in the immediate pre-war years. Since these levels ranged at close to 20 percent for tax collections and expenditures, 10 and 15 percent for outlay expenditures and net value product, one may guess roughly that they would rise to at least 30 percent for taxes and outlays, well over 20 percent for purchase outlays, and close to 20 percent for net value product.[12]

[11] We may expect purchase outlays and net value product to show a relative increase similar to or slightly lower than that suggested for total outlays and taxes. It must not be overlooked that interest on public debt is in the government net value product; and that military expenditures are preponderantly outlay purchases.

[12] This statement applies to the one or two decades immediately following the end of the war. The prospects for a longer post-war period depend upon assumptions concerning longer-range trends in national product and government activity, assumptions that cannot be ventured upon in this paper.

THE NATURE OF THE PROBLEM

The statistical measures of the ratios for the past, as well as the projection of their trends for the post-war decade, are necessarily crude; and as already indicated, such over-all ratios can barely suggest the complex lines of relationship between government and other economic activities in the nation. It would be impossible to discuss adequately here the significance of the trends observed and of their continuation in the future. But a few general comments on the nature of the problem which they raise are perhaps in order.

In considering the relation between government activity and national income, we may frame two questions. Is secular growth in the total product of the economy accelerated or retarded by the increased scope of government activity, as compared with alternative ways of sharing payments and productive resources between the private and the public sectors? Have shorter-term variations in government activity been used and can they be effectively used to reduce undesirable fluctuations in the total performance of the economy?

These questions are of practical significance if we assume that society has a choice in determining the scope of government activity, both for the longer and the shorter run; so that conclusions as to the net effects of government activity can serve to influence the choice in the right direction. They also assume implicitly that maximizing total national product in the long run and stabilizing it in the short run are desirable goals, in terms of which to judge trends and variations in the relative scope of government activity. This second assumption may be denied in favor of some concept of social welfare that goes beyond maximizing a stable national product. Yet, we accept it here as at least a proximate goal.

It would be misleading to state that the questions just framed can be answered in the present state of our knowledge. But we should distinguish two different types of difficulty. The first is that our quantitative knowledge of the processes of interaction between government activity and operations in the private sector of the economy is meager, because only in recent decades has society become aware of the importance of the problem and, therefore, willing to spend

resources to ascertain even the basic economic magnitudes. Even now, there are important gaps in our information that effectively bar objective answers to many factual questions. For example, we don't know even today the income distribution by size among individuals and families; nor do we know the pattern of allocation of income between expenditures and savings at various consumer income levels. Furthermore, since often the collection of important information becomes a matter of social conflict, with considerable pressure against extending the area of economic measurement lest it becomes a basis for centralized government action, it is unlikely that all the necessary data for an adequate study of the interrelations between government activity and national income will ever be secured.

However, progress has been rapid; and the accentuation of problems commonly recognized as of importance to society will lead in the future to further increase of information on the effects, both expansive and retarding, of government activity upon the economy at large. In this measurable part of the problem, there is every reason to expect an extension of factual knowledge and a consequent narrowing of the areas within which conjecture and partial viewpoints can run rampant.

The second difficulty is much more recalcitrant, because it involves finding commonly accepted objective criteria for value judgments. If the analysis of the relation between government activity and activity in the private sector goes beyond the monetary flows into a consideration of *real* costs and returns, we encounter the difficulty of putting government and private activities on an identical value base. That such a base is not provided by the actual rates of pay can be clearly seen by comparing the salaries of the President or the Supreme Court Justices with the compensation of executives or lawyers of equal caliber undertaking similar responsibilities in private business; or, to take the other side of the shield, by comparing the compensation of the holders of political plums or pork-barrel posts with that paid in private business to people of comparable skill and application. The fact is that many activities are in the public sector because it wouldn't be possible to have them carried on within the private sector of the economy on the basis of rules laid down by the market place; just as many activities in our social structure can be carried on within the family but not within the market sector. The actual monetary costs or prices of many government activities are not on the same base as prices in the private sector; and their inclusion in the national product total, as in the tables above, is based upon an arbitrary decision to value government activities by rules that are valid only in the larger and dominating private sector of the economy.

This arbitrary decision is quantitatively and analytically tolerable so long as government activity involves a relatively small proportion of real resources and is producing only a small volume of real goods, no matter how valued. But extension of the scope of government activity means a growing challenge of the validity of continuing to value its real contents at prices actually paid or charged for it—as if these prices were similar to costs and prices on the private markets. A genuine question arises as to whether a new system of valuation is not needed that could effectively translate resources and goods, both in the private and the public sectors, to comparable magnitudes. What that system of values should be, one cannot foresee at present. Perhaps it should relate the performance of both sectors to some notion of social welfare, so that the genuine contribution of private business could be measured while removing some of its dross (such as advertising, patent medicines, etc.); as well as the genuine contribution of the government, again minus some of its waste and inefficiency. Such a commonly agreed upon system of valuation does not exist at present, revealing the absence of an accepted set of social purposes with reference to which both private and public economic activity could be valued. For lack of it, the commonly employed reduction of government and private economic activities to terms of *real* goods is necessarily arbitrary.

This difficulty is clearly distinct from that created by scarcity of data, since it means lack of the very foundation for quantitative measurement. The purpose of stressing it here is to emphasize that the basic questions in the relation of government activity to national income are not answerable in terms of present economic measurement and theory. Evaluation of the significance of these relations requires a set of criteria that transcend the present basis of economic analysis, and is contingent upon formulation of a broader system of social values.

CONFLICT AND HARMONY IN TAXATION

ROY BLOUGH

United States Treasury Department

(Read February 18, 1944, in the Symposium on Taxation and the Social Structure)

TAXATION is an instrument of organized society. Itself a social institution or group of institutions in the broad sense, taxation exists for the service and promotion of other institutions; in and of itself it produces no personal or social utility. Accordingly, the system of taxation rests on the political, economic, and social structure and should be fitted to that structure and adapted to achieving the ends and objectives of the society. Such adaptation does occur, although sometimes belatedly. Thus the general property tax served an undeveloped agrarian America but decayed with the rise to dominance of commerce, industry, and finance. A very current example is presented by Soviet Russia. As long as Soviet Russia followed a policy of equality of incomes, it used the sales tax almost exclusively, but in recent years, when a policy of unequal incomes has been increasingly accepted, the progressive income tax has been adopted.

But taxation is no passive instrument. It exerts an active force of such strength that it inevitably affects the social structure and may importantly alter the course of social evolution. Annual wartime tax collections of 50 billion dollars—virtually as much as the total national income during the depths of the depression—cannot avoid having a major impact on social and economic life. Post-war peacetime taxes will no doubt be substantially lower than wartime taxes, but they can scarcely avoid being so high as to contain an important key to the production and distribution of wealth and income. The effects of such taxation cannot be kept neutral. The potency of taxation is so great that it may be and indeed continually is used as an instrument for making changes which politically dominant groups believe should be made in the social structure. The first Federal revenue act, that of 1789, was for the promotion and protection of industry as well as for the raising of revenue.

Accordingly tax policy is not something to be considered as a separate, independent subject. It is an integral part of general social and economic policy. Likewise the study of tax policy should not be approached from only one side, such as economics. Adequate consideration of tax policy involves, in addition, ethics, law, political science, sociology—indeed every point of view from which the social structure may be studied.

The operations of a tax system involve at least two levels of activity. The first level comprises all the things that happen up to and including the point where the tax is collected from the taxpayer. This includes the passage of laws, their interpretation and testing by the courts, and their administration. The second level of activity comprises the economic reactions to the tax by the taxpayers. These reactions may result in the shifting of the burden from the taxpayer and in a variety of other economic effects. These taxpayer reactions and their repercussions constitute the chief subject matter of the economics of taxation. Governmental policy has relatively little power over taxpayer reactions except as economic institutions and operations in general may be affected by the economic policy of the Government.

To achieve its purposes, tax policy, executed through governmental action, must anticipate and allow for shifting and other economic effects. Wide variations exist in the effects of different taxes, depending on the tax and the circumstances under which it is collected. Accordingly a knowledge of the economics of taxation, as well as of the interrelated intricacies of tax laws, is obviously of primary importance if desired tax results are to be achieved. The translation of a tax policy into a tax program involves what may be called tax engineering. Just as an electrical engineer uses the scientific knowledge at his disposal in the building of an electric motor to certain specifications of size, weight, speed, horsepower, etc., so in the tax field the specialists who are sometimes called "tax experts"[1] face the

[1] The term "tax expert" is generally considered odious by the persons thus designated, but it is too widely used, especially in Congressional circles, to be ignored.

problems of designing measures and devices for achieving the purposes and objectives of tax policy. This process of tax engineering is a vital element in the design of a good tax system. Much of the failure of our tax system to carry out tax policy effectively can be attributed to an unwillingness on the part of policy-makers to allow tax experts an opportunity to perform their functions. In candor, it must be said also that much of the failure is due to bad tax engineering.

Although tax engineering is an important and necessary part of the process of adapting the tax system to changing conditions and points of view, its function is to carry out, not to formulate basic tax policy. The tax engineer can do nothing until he knows the purposes which the policy-makers desire to achieve. This necessary determination of objectvies is not a simple problem, to say the least. In taxation, as in every other field of policy, it is not possible to have all of the good and none of the bad. Tax objectives compete with each other and interfere with each other. An increase in tax equity may involve greater complication for the taxpayers and greater difficulty for administrators. Conversely, the simplification of taxes will almost unquestionably involve some loss of tax equity. Similarly, tax equity may be in conflict with some economic purpose, such as the control of inflation or the promotion of industry. The weighing of these considerations and the choice of a pattern of objectives calculated to produce the maximum benefit is a major problem of policy determination.

But the most important reason for the difficulties of tax policy formation grows out of a basic fact that within the social structure are harmonious interests and conflicting interests. Conflict does not mean—at least not usually—a battle to the death, but rather a situation where one person's gain is another person's loss, while harmony is a situation where one person's gain may also be another person's gain with no one the loser. Conflict and harmony in the social system are analogous in certain respects to the forces operating in the solar system. If the solar system is to exist, there must be forces pushing the heavenly bodies apart and forces pulling them together. The gravitational forces must be stronger than the centrifugal forces if the planets are not to rush madly into empty space. Likewise in the social system both conflict and harmony are inevitable, but harmony must at bottom be the stronger if society is not to be reduced to anarchic chaos.

CONFLICT OF INTERESTS

In taxation conflicting interests are very powerful. Evidence of the existence of tax conflict is inescapable to all who may observe the intense differences of tax opinion, the contradictory tax programs of different groups, and even so small an item as the amount of Washington hotel space preempted by tax lobbyists. To those in position to follow closely the policy-making process, evidences of conflict are often very painfully clear.

The reasons for the strength of conflict in taxation are fairly obvious. Broadly speaking, there is harmony of economic interests in the achievement of maximum production, and there is conflict of economic interests in distributing the product among those who joined forces in the productive process. Taxation is part of the process of distribution—of what is commonly referred to as secondary distribution.

Conflicting interests are stronger in taxation than in most aspects of distribution, because in taxation there is little or no relation between the amount of taxes a person pays and the amount of benefits he may derive from Government. If a person is able through a change in law to transfer part of his burden of taxation to the backs of others, there is no resulting loss to him in the form of a reduction of governmental services. Moreover, the tax bargain, if it may be so designated, is in the political field, and there are no economic acts of the individual whereby he can affect the amount of his burdens by withholding goods or services from the market. This applies, of course, to the first impact of the tax; he may to some extent be able to affect the subsequent shifting and incidence of the tax; but it is in the political field that the conflict principally appears.

Major conflicts of interest over tax policy arise primarily because persons are in different economic positions or different geographical jurisdictions. Perhaps the most important economic difference is that based on the volume of wealth or income, which is reflected in the conflict of interests over the progressiveness of taxation. The change in three decades from the regressive Federal taxation of pre-income-tax days to the highly progressive income taxes of today was highlighted by many struggles.

While the emphasis in tax discussions has been largely on differences in the amount of wealth or income, other economic differences also give rise to conflicts of interest over tax policy. Thus,

there are differences in the source of income: a person may derive income from property alone or from personal effort alone, or from various proportions of property and effort. There are differences in the occupation and the industry from which income is derived. Differences are likewise present in the forms of wealth, especially between tangible and intangible property. Other differences concern the disposition of income between consumption and savings, among various kinds of investments, and among various types of expenditure.

Conflicts of interest in taxation arise also from geographical differences, that is, differences in the political jurisdictions in which people live. Responsibility for financing the activities of government is divided among some 175,000 jurisdictions. All persons living within any one political jurisdiction have some interests in common which conflict with the interests of persons in other jurisdictions.

An individual is not limited to only one interest. He always has at the same time several kinds of economic and geographical interests, each in common with various other individuals. Thus a manufacturer who has an additional income from corporate stocks and owns his own home has certain interests in common with real property owners in general, certain other interests in common with manufacturers, others with owners of securities, others with the corporation in which the stocks represent equity, and still others with consumers in general. He also has interests in common with other persons in the same political jurisdictions—city, county, and state—regardless of the relative economic situation of those persons. His interests and outlook are thus not unitary; he is drawn in opposite directions by his competing interests. In taking an intelligently selfish stand on questions of tax policy, his position will depend on the evaluation that he makes of his competing interests. Relative values of these interests to him may change from time to time, thus changing his attitude toward tax policies.

This conflict within each individual prevents the tax struggle from being a clear-cut fight with every person permanently lined up on one side or another. In many kinds of tax questions, the balance of interests will be so close that a large number of individuals will be relatively neutral. This group of neutrals often holds the balance of power between conflicting groups and thus is for practical purposes the public. Although the true public is commonly at least a majority of the people, the effective public in any tax conflict is this group of relatively neutral persons. The composition of the effective public group may change with each succeeding tax issue.

Although no two persons are likely to have the same pattern of economic and geographical interests except by accident, there are degrees of similarity and contrast. Since political power largely determines tax policy, persons with like interests tend to form groups for the more effective expression of their opinions. These groups have varying degrees of effectiveness in dealing with the makers of tax policy, but in general have a great advantage over persons not working through group organizations. On the geographical side, unity and conflict of interest are formally recognized in the system of representative government under which representation is based on geographical districts.

HARMONY OF INTERESTS

The spectacular nature of conflicting interests in taxation tends to obscure the presence of harmony of interests. Some harmonies are direct and obvious. Nearly everyone prefers certainty to uncertainty and confusion in the amount of his taxes. Nearly everyone is benefited by a low cost of tax administration and low cost of compliance.

Other harmonies underlie apparent conflict or at any rate set limits to the conflict. For example, a high protective tariff on manufactured products might so destroy the foreign markets of domestic farmers and diminish their purchasing power as to affect adversely the markets of manufacturers with the result that manufacturers would really be better off with lower tariffs. In this case there is an area of conflict between farmers and manufacturers, but beyond a certain point the conflict may be replaced by harmony, since both may benefit from the same policy.

Another type of basic harmony is inherent in the concept of tax justice. Although what is just is a matter of opinion, there is a good deal of agreement over large areas. Justice is a moral end or objective, and its violation outrages the moral sense of people. Moreover, good will among various groups in society is necessary to the effective functioning of economic organization. Flagrant injustice in taxation destroys good will and leads to destructive antagonisms. Taxes imposed by the political power of one group may so destroy the good will of other

groups that the harm from the resulting non-co-operation more than wipes out the advantage to the apparent beneficiary. Some superficial advantages that an individual or group might obtain by political force always must be sacrificed in order that there may be peaceful and orderly political, social, and economic relations through which greater advantage for all may be secured.

Still another kind of harmony is based on the interest which the members of a group seek not each for himself but each for the whole group. Such group loyalties or common interests usually arise when the group is in conflict or in competition with other groups, and the loyalties tend to disappear in the absence of such conflicts or competition. For the nation as a whole this kind of harmony in taxation has proved difficult to achieve even during time of war when national unity is greatest.

CONFLICT, HARMONY, AND TAX ISSUES

Problems of forming tax policy take the form of tax issues. A tax issue involves choosing among alternative courses of action with regard to any of a large variety of tax questions such as the total amount of taxes, the types of taxes to be employed, the relative emphasis on each type, rate schedules, and exemptions, as well as percentage depletion, basis for gain and loss, invested capital credit, contemplation of death, pay-as-you-go, floor-stocks tax, definition of proof gallon, powers of appointment, withholding certificate, post-war credit for the Victory tax, carryback of losses, and many other technical matters.

The absence of a tax issue in a certain area does not necessarily mean that no conflict of interest is present. Tax issues do not ordinarily arise if the interests of a very small minority are opposed to those of a large majority. In such a case, the decision is taken for granted and no problem is recognized. The minority feels its helplessness and does not even bother to raise the issue. The possibility of favorable action may be so remote that they do not even think about or recognize their own interests. Issues come to the foreground only when opposing groups are of substantial size.

Tax conflicts may present themselves in various forms. Two group interests may be pitted against each other, or one group interest may be pitted against the combined opposition of several other groups, or one group may be struggling against the inertia and passive resistance of the remaining groups. There need not necessarily be an apparent, direct clash between groups. The issue is perhaps more likely to take the form of one group demanding that its taxes be lowered or not raised, as the case may be. Representatives of such a group commonly say they have no desire to increase the tax burdens of anyone else. This is no doubt often true. Nevertheless the tax burdens on others must be raised, eventually if not immediately, to the extent that the burdens of the particular group are lowered. This conclusion rests on simple arithmetic, but if it were appreciated fully, there would be fewer cases of especially favorable treatment in the tax law.

The apparent tax issue is not necessarily the real one. For example, the total amount of taxes to be raised is an important issue which often conceals more basic issues. Thus, high taxes are often opposed when the real issue is the volume of governmental expenditures. This appears to have been an important factor in the attitude of the Congress towards the recent proposals for more than 10 billion dollars of new taxes. The issue of the volume of taxes may also conceal basic conflicts of interests over such apparently extraneous matters as the desirability of inflationary price rises or other economic effects which might come from higher or lower taxes.

The issue of the proper use of the taxing power also commonly conceals some other issue. An examination of the circumstances under which various groups oppose the use of the taxing power for other than revenue purposes makes one justifiably suspicious of arguments in this field. Many such opponents are undoubtedly sincere and not motivated by narrow self-interest. Too frequently, however, the real objection to the use of the taxing power for some social or economic purpose is that such use would be contrary to the objector's self-interest. The conflict of interest over the social or economic result of tax policy is concealed under the guise of an objection on grounds of principle, the principle being that the taxing power should not be used for any purpose other than revenue. When the shoe is on the other foot, however, and the use of the taxing power can be made to bring advantage to their group, little difficulty is found by such persons in supporting the measure. Highly technical provisions of the tax laws may constitute important issues in this respect, because through such provisions, which may be entirely incomprehensible to the general public, special benefits in

form of special low rates or exemptions are frequently extended to special-interest groups.

Tax issues also arise where there is underlying harmony. In this case the problem is a soluble one, at least in theory, depending for solution only on an understanding of the facts and on the application of proper tax engineering techniques. It is soluble because there is a one best answer to the problem. In case a tax issue is based on conflict, however, the issue is essentially insoluble. There is no one demonstrably best answer, because a good answer to one person may be a very bad answer to another.

THE POLITICAL FUNCTION AND THE PUBLIC INTEREST

Still, there must be an answer. It is this necessity for an answer that makes political action, that is, policy-making, a vital function. The function of the policy-makers is to decide among the conflicting interests with respect to any particular tax issue.

The test of good tax policy is whether it is in the public interest. What is in the public interest is good, what is against the public interest is bad. Where the interests of all of the people are in harmony the public interest is easy to determine, since it is the common interest of all. Where there are conflicts of interest, however, the public interest is not so easy to determine. Yet the public interest must somehow be located in the maze of conflicting private interests. Where there is a clear majority of people on one side or the other, the public interest will ordinarily correspond with the interests of the majority. But commonly there is no clear-cut majority one way or the other. There is usually a relatively neutral or indifferent group which decides the issue and with respect to the issue represents the public. The part played by the effective public group is to evaluate the claim of each directly interested group and to decide the issue in the manner that will best serve the interests of the members of the effective public group or what they conceive the public interest to be. Since this group is relatively neutral, it is in position to decide the issue fairly and with considerable objectivity. Each directly interested group endeavors to persuade the neutral public group. The decision may be in favor of one of the directly interested groups or it may be a compromise between their demands. As the result of the action of the public group an equilibrium is reached which will last until changing circumstances or changing opinions again raise the same or a related issue.

This choosing and compromising is the main function of the politician in the best sense of the word. It is not a function which can be performed by tax experts or tax engineers. It represents at bottom a determination of what is in the public interest. Only the public is in position authoritatively to state its interest. The people are often sadly in error as to what are their interests, but obviously no one person can be sure he knows what is the public interest.

It will be readily agreed, no doubt, that the private interest groups are not likely to be very much interested in tax principles. They are interested in getting what they want; the principles they espouse are likely to fit their interests as they see them. The neutral public group, however, is open to persuasion on the basis of principles. The interested groups often seriously abuse tax principles in urging upon the neutral public group the view that their interests are in harmony with principles accepted by the public group. For example, many crimes of tax propaganda are committed in the name of the principle of ability to pay.

TOWARD "REASONABLE" TAXATION

Because of conflicts of interest and the part they play in tax issues we cannot hope for ideal taxation. Even if there be such a thing as ideal taxation in contemplation, no one is in position to demonstrate what an ideal system is, and the conflicting groups in the community certainly would not recognize it as such. What we can hope for is "reasonable taxation."[1] Reasonable taxation is taxation which constitutes at any given time the nearest practicable approach to the public interest. Reasonable taxes cannot be made fixed and unchanging; they will not stay reasonable with the passage of time and changing conditions. People who lose today in their attempt to get tax changes which would promote their interests will reopen the matter another day when the circumstances seem more favorable. Perhaps on that occasion their private interests will be judged to coincide with public interest and the desired tax change will be made.

---

[1] *Cf.* John R. Commons, *Institutional Economics* (New York, 1934): 805–840 (particularly 820–821). The present writer is deeply indebted to the classroom teaching and published works of Professor Commons with respect to the ideas discussed in this paper.

This importance of tax conflicts and the prospect of an ever-changing tax system are very distressing to many people, which is not surprising. Conflicts are confusing and leave one without a nice solution. Conflicts tend to take the form of log-rolling for special privileges and loopholes, with a consequent erosion of the tax base and degeneration of the tax system. A long catalogue of failings in the tax system can be laid at the door of conflicting interests. The political process has frequently been distorted to serve narrow special interests which by no stretch of the imagination could be considered the public interest.

One reaction to the problem is to minimize and ignore conflict. This, by and large, is the aproach of treatises and textbooks on economicss and public finance. Perhaps it is thought that conflict is like sin, something to be kept down by keeping the mind on the higher harmonies. More seriously, a tax system based on high principles reflecting common interests and harmony is both intellectually and emotionally more satisfying than one based on conflict. Tax theory based on harmony can be much more nearly "scientific." Moreover, it is through unity that society is strengthened and through conflict that it is disrupted, sometimes to the point where totalitarianism seems the only solution.

But there is no use in glossing over the facts. Conflict will not go away simply by being ignored; the only result of ignoring it is that the evils growing out of conflict will continue to flourish, the tax system will continue to be condemned, and the politicians will continue to be castigated. The conflict is there anyway, and in many cases private self-interest is going to decide tax issues. To recognize the conflict and bring it into the open would seem the better course. Then one can see better whether the public interest is really being achieved from conflicting private interests. After all, if each person evaluates his own competing interests intelligently, if each group really knows its own best interest, if the effective public group understands the relations between group interests and the public interest, and if the political machinery fairly and effectively represents the different interests, the result should be the achievement of public interest in all cases. This is an imposing list of "ifs." They may come true but only in "never-never land." They are far from being true in the United States today. The point is that the future improvement of tax policy and the tax system rests largely on making them increasingly true.

Unfortunately, there is no secret formula for improving the tax system; there is nothing new on which to pin hope. The old standbys are about all there are, namely, increased understanding, wider dissemination of knowledge, and the improvement of the political machinery. So far as the last of these is concerned, the deficiencies of the system of representative government in the United States have been pointed out many times. While there are no doubt special lines of revision which might be mentioned with particular respect to taxes, time does not permit the discussion of these possibilities in this paper.

The considerations underlying tax policy may be said to look in two directions, the past and the future. The principle of tax justice looks to the past and endeavors to levy tax burdens in accordance with the ability of people to pay such taxes as measured by past evidence. It is believed that makers of tax policy, however, are usually more concerned about the effects taxes will have on employment, production, and the condition of the economy generally than they are on ability to pay, although the latter is not disregarded. They are, at their best, interested in levying taxes in such a way as to contribute to the greatest economic well-being in the future. The generalizing of this point of view into a principle of taxation would mean that taxes would be kept relatively low at those points where taxes interfere with production, and made relatively high at those points where such taxation would not interfere with production. Putting it another way, taxes would in so far as possible be made to fall on social surpluses instead of on social costs. The development of such a principle to the point where specific application could be made to the different tax issues as they arise for settlement would provide an invaluable tool for the policy-maker. The development of such a principle depends in part on a much greater knowledge of the effects of taxes than exists today. This is true even though knowledge has outrun practice in some respects.

There are many sectors of the tax field in which further study, specific quantitative study as distinguished from general analysis, is necessary if the jobs of tax policy and tax engineering are to be done adequately. There needs to be a further advancement of knowledge concerning the extent and limits of conflicts and harmonies of interest. It is almost certain that the area of harmony is

much larger than ordinarily appears to be the case. Continued study should narrow the range of conflict and reduce its intensity. This is very important if the more unfortunate effects of conflict of interest are to be minimized. Further study is needed also to determine the relationship of taxes to production, employment, and other socially desirable economic factors.

Along with this expanded economic research should go a more widespread public dissemination of tax information. There seems to be a firm belief by most people, even very competent people, that they will never be able to understand taxes. This is only an inferiority complex, but their problem is not clarified by the many conflicting programs which are pressed upon them, and the frequent wide divergence of opinion among economists has not helped the situation.

Fortunately, not everyone has to be fully informed. We are an organized people and rely largely on the leaders of our organizations. These leaders need to absorb as much as possible of the type of information which will be forthcoming from economic research in taxation. Education of this kind will be helpful in distinguishing long-run interest from short-run interest and in recognizing private self-interest which poses as public interest. Such education is necessary also to lead to the broader harmony which underlies much of tax conflict and which, if promoted, may lead to unity rather than divergence.

It is suggested that the encouragement of research and the education of the leaders of the community call for the establishment of a journal of public finance. There is no journal devoted to the whole field of public finance and that field exclusively. Such a journal could make available to the public in relatively non-technical form the results of research in taxation economics. Such a journal could also cut across the boundary lines between economics, political science, law, and administration—a necessary procedure since most tax matters naturally cut across these lines.

Great care would have to be taken to see that such a publication did not become a forum for special interest groups to urge policies which they desired. Indeed, it is believed that in such a journal policy recommendations should be kept to a minimum, with the major emphasis on research.

There may be other better ways of achieving greater knowledge of the effects of taxation and greater dissemination of this knowledge. Certainly some changes are needed in American tax policy formation. At the risk of repetition it is to be emphasized that the most disquieting aspect of taxation is the dominating role of conflicting interests based too often on shortsightedness and ignorance. The tax system shows signs of degenerating into a collection of myriad provisions conferring special privileges on numerous minority interests. Conflict is inevitable and should be accepted as a necessary and desirable part of the process of arriving at a tax policy which will be in the public interest, but when conflicts disrupt the social and economic structure, society is likely to create artificial uniting forces and institutions, and these may be very dangerous to a democratic way of life.

The problem is to find ways to use tax conflicts to produce a finer tax system, one more attuned to the ever-developing needs and ever-changing emphases of the social structure.

THE SALES TAX

ALFRED G. BUEHLER

University of Pennsylvania

(Read February 18, 1944, in the Symposium on Taxation and the Social Structure)

THE social importance of government financing has grown with the phenomenal expansion of the federal budget to 100 billion dollars annually under the impact of World War II. After the war we may come to regard a federal budget of 20 billion dollars or more a normal event. The social implications of this transformation are not yet clear, and groups of citizens are meeting here and there over the country, as the American Philosophical Society is meeting here today, to discuss the problems accompanying the transformation. One of these problems is the formation of a tax program adequate for the huge requirements of war and post-war financing.

Tax systems are always the outcome of political and social struggles among groups and classes, each seeking to escape as many of the unpleasant consequences of taxation as possible. The tax system, at best, is an imperfect mechanism of revenue and fiscal control, because government financing has not yet developed into a science.

The slogan, "taxation for revenue only," has been popular among those who like to imagine that taxation can be imposed in such a manner that existing social institutions will not be disturbed. But we know now that every tax carries with it social repercussions, direct and indirect, immediate and remote, and that the purpose of government financing is not to leave things as they are but to provide some new service or to discharge a function more effectively through public action. Many persons, indeed, who ask for taxation for revenue only uphold the use of taxation for certain favored objectives, such as the protection of industry from foreign competition, the exemption of new manufacturing industries from taxation, and the broadening of the tax base to make the masses tax-conscious.

Despite all of the criticism of the use of taxation for social control and its deficiencies as an instrument of reform, the demands for taxes to curb inflation in wartime and for reforms in taxation to provide incentives to work, save, invest, and promote enterprises after the war are increasing. The tax student may concede that if a given tax is the most effective of the available alternatives for the furthering of clearly desirable social objectives, its use for social regulation may be accepted. In view of the growing recognition of the possible effects of taxation upon social institutions, it is necessary to examine the sales tax as a tool of social as well as fiscal policy.

OBJECTIVES OF WAR AND POST-WAR FINANCING

The objectives of the war economy appear to be (1) the maximization of production, especially for war needs, (2) the diversion of all necessary resources to the war effort, (3) a just sharing of the available civilian goods and services, (4) the equitable distribution of the costs of the war, and (5) the preservation of the stability of our society both during and after the war. The raising of revenue thus becomes incidental to the winning of the war and the maintenance of the American way of life. But government spending and borrowing cannot be continued without ample tax resources, and the raising of revenue is a necessary function of the war economy.

It appears to be generally agreed that our economic objectives after the war must be (1) the greatest possible output of goods and services for civilian consumption, (2) the just distribution of the outpouring of material thus produced, (3) the avoidance of mass unemployment and depression, (4) the achievement of economic prosperity without economic instability, and (5) the equitable sharing of the costs of government by carefully planned taxation, which will be adequate for the support of necessary government functions, but will not deaden enterprise and slow down the national production.

In looking ahead to the fiscal problems of the remaining months or years of the war and the years which will follow, what role should be assigned to the sales tax? In attempting to

answer this question, the tax should be weighed in relation to the tax system as a whole, and not as an isolated measure. What revenue and social effects would be produced by the federal tax system if a sales tax were introduced? So far as our time will permit, let us examine some of the possible revenue and social implications of the sales tax, a form of revenue which has been the subject of long and heated controversy in this country.

AMERICAN SALES TAXES

After World War I, when the excess profits tax and the high surtaxes of the individual income tax were much criticized, Congress was besieged with requests for a general sales tax. This revenue had been known in only a few areas before the war, but it soon spread into many countries during the world-wide search for new revenues to meet the requirements of post-war financing. It was clearly a by-product of World War I. The pressure against a general sales tax was strong, however, and Congress refused to introduce it. The demands for a general sales tax were revived during the depression of the 1930's, when income tax collections slumped and larger tax revenues were desired, and again the tax was rejected.

Meanwhile the states were adopting the general sales tax or, as it is often called, the sales tax, either as a tax upon retailing or as a tax upon retailing and other types of business. A few states had experimented with the sales tax during the nineteenth century, but the rush to this form of revenue began after World War I and was renewed during the depression of the 1930's, when traditional revenues failed. Some of the state taxes were abandoned while new taxes were being added. In 1943 there were 23 states employing sales taxes, from which they collected an estimated $665,000,000.

The experience with the state sales taxes, which were often adopted as emergency revenues and later retained for normal requirements, has proved that the sales tax is practicable, but not easy, of administration, that it burdens the lower income groups more heavily than the upper because the former spend relatively more of their incomes, that the attempt to tax commodities coming in across state boundaries creates difficult problems, and that substantial revenues may be enjoyed. The unique American contribution in the field of sales taxation has been the retail sales tax, which would probably be the most suitable type of sales taxation for war needs and also for a peacetime economy in this country.

TYPES OF SALES TAXES

A federal turn-over tax, levied on every transaction from the producers to the consumers, would provide large revenues at a low rate and has often been proposed. Such a tax would encourage the integration of business processes, however, because the tax could be reduced by combining productive and distributive processes under one management, and it would give strong competitive advantages to the integrated industries over their non-integrated rivals. The tax would also be difficult to administer because it would apply to all types of business. If it were adopted in wartime, after price ceilings had been established and wages and farm incomes had been stabilized, it would upset costs and prices all along the line and disorganize the established controls.

A tax upon the sales of manufacturers only would offer administrative advantages, but its base would be relatively narrow and a high rate of tax would be required for great revenue. This tax would also change costs and prices along the line and would play havoc with price ceilings and stabilized incomes. The tax would not be so apparent to consumers as a tax upon retailing, and difficulties would be faced in finding satisfactory definitions of manufacturing. As a wartime revenue measure and a device to drain off excessive consumer funds, a manufacturers' sales tax would be less attractive than a retail sales tax. In peacetime it is possible that the country would prefer a tax upon manufacturing, which Canadian experience has demonstrated may be productive and practicable.

A tax upon wholesaling would have some of the advantages and disadvantages of the tax upon manufacturing. The tax base would be difficult to define, and costs and prices would be disturbed. The tax base would be narrower than that of the retail sales tax, and a higher rate would be necessary for an equivalent revenue. The tax upon wholesaling is not so popular in this country as the tax upon retailing. It has been introduced in Great Britain, however, and it is not without advantages.

The retail sales tax would probably be more popular, or less unpopular, in this country among business men and consumers than another type of sales tax. The state taxes upon retailing

have become well known and they have many supporters. In wartime, the retail sales tax would be more suitable for revenue and inflation control than a tax upon manufacturing or upon wholesaling, or a turn-over tax. Because of its broad base a relatively low rate would yield large revenues. The Treasury has estimated that a 10 percent tax in the calendar year 1944, without exempting food, would return over 6 billion dollars, and a 5 percent tax over 3 billion dollars. If all food sales were exempt, the yield of the 10 percent tax would be cut by over 2 billion dollars. These sums are important, but they would not provide major revenues in a federal tax system which is producing over 40 billion dollars annually.

A retail sales tax would disturb costs, parity prices, and price ceilings less than another variety of sales tax. It would not be pyramided along from producer to consumer, like a turn-over or manufacturing tax. The tax would be collected by retailers from consumers, to whom it should be visible, and it should remind them of the costs of war financing. Consumer consciousness of the tax might assure its repeal after the urgency of war requirements had passed, which many would consider an advantage, especially if they oppose the sales tax as a normal revenue.

At a cost of some revenue, the retail sales tax could be made more equitable by exempting food, the most expensive item in the budgets of the lowest incomes. The tax would not fall upon rentals or payments for living quarters, and it would presumably exempt medical, dental, funeral, and other services. If it were desired to tax certain services, such as those provided by barber and beauty shops, electricity and gas for home consumption, and amusements, they could be taxed with supplementary excises. In the application of a federal sales tax, it would probably be desirable, for administrative and other reasons, to confine the tax to sales of tangible personal property. Thus a large part of the family budget would be exempt, and the exemptions could be extended, with a loss of revenue, if social policy required this step.

The retail sales tax could be employed, along with the excises, to supplement the progressive income, estate, and gift taxes and to reach persons with spendable funds in excess of subsistence requirements who escape the progressive taxes. The tax would be collected conveniently with daily purchases, and the consumers liable for the tax would be current in their tax payments as soon as they paid for their purchases.

Because of its advantages, if a federal sales tax were adopted, it would probably be imposed at the retail level in wartime, and the popular preference for the retail tax would probably assure its use after the war, if a sales tax should be introduced at that time. For these reasons my remarks will be directed primarily to the problems of the retail sales tax, and I shall assume that the tax would be confined to sales of tangible personal property.

THE SALES TAX AND INFLATION

The retail sales tax has recently been proposed by many persons and groups as a tax admirably suited for war conditions, when consumer spending should be curbed and as large a part of the federal expenditures as possible should be met out of taxation rather than borrowing. A retail sales tax might siphon out of consumers' purchases some billions of dollars which might otherwise flow, in a large measure, into the retail markets and bid up the prices of scarce commodities. The point has been repeatedly emphasized by Government spokesmen that consumers have been receiving many billions of dollars of income in excess of the value of available goods and services, and that after allowances are made for debt payments and various investments, an inflationary gap of billions of dollars will remain in the form of funds which may be hoarded or spent at the will of the consumer. The excess of spendable funds should be immobilized by taxation or a program of savings, if they are not to constitute a constant inflationary threat which may at any time be released in the retail markets.

As it is ordinarily collected, the retail sales tax could not be personalized in a manner to hit only the excessive purchasing power of those individuals with increased incomes or others who are eager to spend their funds on luxuries or non-essentials produced in competition with war materials or on commodities needed in the war program. With the exemption of food, rentals, and other necessary expenditures, however, the tax would concentrate largely on the less necessary spending, but not at rates graduated according to the amount of spending. In a rough impersonal way, the tax would act to curb or penalize spending. It would not, however, take the place of priorities, rationing, and price ceilings, which are essential for the selective control of particular lines of consumption and

production and for the equitable sharing of civilian goods.

By discouraging spending and rewarding saving, the tax could supplement other inflationary controls. The incentives to work harder and longer would probably be dulled less by a sales tax than by steep increases in the income tax, which would take a heavy toll out of every dollar earned, whether it was spent or saved. The individual could pay his debts and set aside his savings with the knowledge that these uses of funds would not be taxed. If spending were postponed after the war, it would probably pay no higher, or even a lower, rate of sales tax.

The tax would fall upon the spending of hoarded funds and other savings as well as upon spending from income sources. It would also reach the spending of other funds from non-income sources, including the interest on tax-exempt securities. While the tax would not encourage the investment of savings in war bonds or productive investments rather than the hoarding of funds, it would strike at the spending of cash hoards. The liability for the sales tax might be met in some measure by sacrificing saving rather than spending, but in general it would probably lessen or penalize spending rather than saving. Thus, in a rough impersonal way, a federal retail sales tax could be directed against consumer spending with some advantages in buttressing the other controls over inflation.

In its opposition to the retail sales tax the Treasury has insisted that the tax would not only act in a measure to curb inflation, but that it would also have serious inflationary effects because it would operate in several ways to raise costs and prices. First, it would be impracticable to confine the tax exclusively to the uses of goods for consumption. If fuel, tools, equipment, and supplies employed in business were taxed, the taxes would enter into costs and prices. If exemption were attempted for all business uses of commodities, the exemption could be only partially successful at best because some commodities may be used for either business or non-business purposes, and the experience of the states reveals the great difficulties involved in attempting to confine a tax to retail sales to consumers.

The tax would modify the costs of agricultural production, as it fell upon the purchases of commodities utilized for production by the farmers, and parity prices would have to be altered, with ensuing changes in wholesale and retail prices.

A retail sales tax would also raise the living costs of workers and farmers and incite demands for higher wages and incomes, especially if subsistence expenditures were not exempt. If wages and farm incomes were increased, business costs would advance and prices would tend to rise. The opposition of organized labor to sales taxes is well known, and the farmers might join the workers in demands for concessions to offset the additions of sales taxes to retail prices.

To the extent that a retail sales tax was shifted forward to consumers it would raise living costs; and wage, farm income, and price upheavals might follow, the Treasury argues.

These arguments are weighty and should not be dismissed lightly. In order to minimize the indirect inflationary effects of a retail sales tax, it should be introduced before war inflation sets in, and the readjustments in prices can take place without vociferous demands from highly organized groups of farmers and workers for higher rewards. Moreover, the confusion resulting from changes in price ceilings which are produced by a sales tax can be avoided for the most part if the tax is introduced before price-fixing operates. Once inflation has set in and labor and farm organizations are pressing the Government for concessions to offset advancing prices, it will be especially difficult to hold the line against increases in wages and farm incomes.

The immediate and the long-run effects of a sales tax on prices should be distinguished. Individual prices will rise or fall as a result of the tax, depending upon changes in costs, profits, prices, and other factors, as the processes of tax-shifting operate. If the tax were very heavy, it would act in time to depress industry and prices in many instances. The long-run effects will be unequal and they will depend upon the interplay of variable forces.

The effects of the tax upon Government financing should also be observed. The immediate rises in prices which resulted from higher costs would increase the sums paid out on Government purchases. On the spending side, the effects of turning over tax collections into contracts for materials, services, and other items should also be noted. Funds would be transferred from the taxpayers to the Government, which would transfer them to those furnishing it with goods and services.

The collection of the tax at the retail level would help to prevent its inclusion in the costs of articles going into business uses, both because the

tax would be collected directly from consumers and also because it is generally true that purchases at retail are made by consumers rather than by business concerns. So far as possible, commodities used in business should be exempt, but, as already stated, this is not completely practicable.

The argument that a retail sales tax would have inflationary effects because it would be regarded as a factor raising living costs is applicable to other taxes as well. One may say at the outset that the workers and farmers are pressing to increase their incomes and that one excuse or another will be given in requesting higher rewards. Any new tax of broad application and considerable weight may be employed as a lever to move the Government to grant higher incomes to politically powerful groups. The effects of a retail sales, income, spendings, or other tax program upon wages and farm incomes will therefore depend upon the willingness and ability of the Government to resist pleas that tax sacrifices be wiped out by concessions.

Last autumn, when the Treasury proposed higher personal income taxes, even though much of the increase was to be refunded after the war, President Murray of the Congress of Industrial Organizations served notice to the Ways and Means Committee that his labor organization would regard the enactment of the Treasury proposal as an act unfriendly to labor, and that it would serve as a signal for an immediate campaign to compel the Government to authorize higher wage scales. The question at bottom is this: What groups must pay for the war? The groups which resist a retail sales tax may also oppose a spendings tax or higher income taxes unless these measures entirely exempt the lower incomes.

THE REGRESSIVITY OF THE SALES TAX

A powerful popular argument against a federal sales tax has been the often repeated charge that such a tax would be regressive in its burdens on consumers, to the extent that it could be shifted to them. That is, the tax would press most heavily on the lower incomes because of the general principle that the part of the income spent varies inversely with the amount of the income. The tendency to spend will also depend upon the number of dependents, and they are usually more numerous in the lower income families.

Other inequalities will also arise. Farm families produce much of their food and usually supply more of their needs than urban families. Even if food sales are not exempt from taxation, the farmers will escape a considerable part of a retail sales tax. The exemption of rentals and payments for homes will also bring unequal advantages, depending upon the importance of these items to particular families.

To lessen the regressivity of a retail sales tax and its inequalities of burden, the exemption of foods would be helpful and would probably be required because of popular pressure, unless the tax rate were very low. Since the tax would probably not apply to rentals and payments for homes and necessary medical and other services, much of its possible regression in burden would be removed.

Another type of exemption is sometimes proposed, the exemption of a certain amount of purchasing. This might be $200 for adults and $100 for children. The exemption would be given through the issue of coupons which would entitle the holders to purchase designated amounts at retail free from tax. At the time of purchase, a coupon would be relinquished to the retailer. The exemption could be given to the total population or it could be limited to the smallest incomes. The plan would offer administrative difficulties and opportunities for tax evasion by trading coupons back and forth between taxpayers willing to sell coupons and those wanting to buy, or by the free exchange of coupons, where it was found that the exemptions were larger than necessary for some persons. Because the purchasing power of money varies in different sections of the country, a flat exemption would give greater advantages in some areas than in others.

In weighing the social effects of a sales tax, on the assumption that it is shifted to consumers, the effects of other taxes in the revenue system should also be studied. The federal, income, profits, estate, and gift taxes, which together provide the great bulk of the revenue, are levied at high rates and with progressive effects upon the population, excepting the low incomes, which are very largely exempt. The excises do fall upon articles of mass consumption with regressive effects, but they provide only a minor part of the total revenue, and some of the less necessary lines of consumption are taxed at relatively high rates, with some progressive effects. All in all, the progressive taxes provide a very substantial counterbalance for the excises with regressive

burdens. If a sales tax were introduced, more emphasis would be placed upon the relatively regressive taxes, but it would be possible, if it were held to be desirable, to increase the progressive taxes somewhat.

To some observers the regressive character of a sales tax appears to be an important advantage because the tax can be employed to place a greater part of the cost of the war upon the workers and others with moderate incomes. It also reaches those with surpluses of spendable funds who may pay little or no income tax, including the workers and the farmers. The income tax, with its exemptions and deductions, can be personalized far more than a sales tax, and most Americans would probably prefer to have the income tax extended to its limits before a sales tax is introduced. But as the income tax rates rise, those with debts and contractual commitments are hit harder and harder, and as the exemptions drop, the administrative problems of tax collection increase. When the practicable limits of the income tax are approached, other taxes must be employed to raise revenue and to check inflationary spending. Here the excises and the sales tax become available.

The Treasury, which has repeatedly denounced the regressivity of the sales tax, advocated a flat-rate 10 percent spendings tax in 1942 as one feature of a program with offsetting progressive features. The flat-rate tax would be applied to all consumers, except those with only enough funds to pay for a bare subsistence. It was to be supplemented by a graduated spendings surtax. The first tax would plainly be regressive in its burdens. It was, in effect, a 10 percent sales tax which would be collected from individual consumers in conjunction with the income tax. Its burdens and inflationary tendencies would seem to be similar, in many respects, to those of a retail sales tax. The flat-rate spendings tax was to be refunded after the war, but a sales tax could also be refunded if a plan similar to that proposed by Senator Danaher were adopted and the consumer were required to buy defense stamps with all of his purchases at a designated ratio, perhaps at a rate of 10 percent, of purchases.

Congress rejected the Treasury spendings taxes because they seemed to be highly complicated and confiscatory in rate. A simpler method of accompanying a flat-rate consumption tax with an additional graduated tax could have been found in the levy of an additional income tax at graduated rates upon spending out of income, that is, upon net income, after subtracting deductions for debt payments, war bond purchases, and savings out of income. This would have obviated many of the difficulties arising in the attempt to tax all spending from non-income as well as from income sources at graduated rates by the spendings surtax.

The retail sales tax is a quite adaptable instrument of fiscal policy if the country is willing to accept some frills in collecting it. It can supplement progressive taxes, or it can be converted into a graduated consumption tax with the help of stamps or coupons which would be issued to consumers. The stamps could be issued in such a way that the tax rates would increase as purchasing increased, as shown by the numbers of stamps turned over to retailers when purchases were made. Or graduation, within moderate limits, could be accomplished by refunding the tax after the war in such a manner that the lower spending brackets would be given larger refunds than the upper. This plan would also require the use of stamps, which would be issued by retailers with purchases and would be saved for refunds of taxes at an appropriate time.

The nuisance of issuing, handling, and checking stamps is a serious practical obstacle in wartime, when rationing coupons are numerous, to the introduction of a graduated-rate sales tax. In peacetime, if the country were willing to accept the red tape, a graduated sales tax would be practicable, perhaps, but it would entail serious problems of tax administration and tax compliance.

I have spoken of the objections to the regressive burdens of the sales tax and the possibilities of correcting or counterbalancing them by various devices. The effects of the sales tax upon the entire community as well as the effects upon the taxpayers should be weighed, and if it is found in wartime that the menace of inflation calls for drastic remedies, the sales tax is available, along with the excises, to drain off spendable funds from the retail markets which are not absorbed by the income tax.

In peacetime a retail sales tax would offer advantages in curbing consumer spending and in paying off governmental debt during boom periods. During a depression it would place burdens upon consumer spending at a time when it should be encouraged if economic recovery is not to be restrained. For this reason, many

persons who advocate a sales tax during war and post-war prosperity oppose its use in times of depression.

THE ADMINISTRATION OF THE SALES TAX

The experience of foreign governments and our states indicates that a federal retail sales tax should be administratively feasible, but that many difficulties would have to be overcome. Taxable sales would be hard to define, and the accounts of over two million retailers and their tax returns would have to be checked carefully to prevent tax evasion and to maintain the revenue yield. The higher the rate of the tax, the more would be the incentive to tax evasion. Secretary Morgenthau recently told the Senate Finance Committee that a retail sales tax could be collected at a cost of 18 million dollars and with 6,700 additional employees in the Treasury. This cost, it should be noted, would be relatively moderate for a tax yielding 3 billion dollars or more.

The states appear to have found the retail sales tax about as difficult as the income tax to administer. The difficulties they face in trying to tax commodities crossing state boundaries would not be encountered in a uniform federal sales tax, however, which would apply equally to interstate and intrastate commerce.

Like the income tax, the spendings tax, and other important revenues, the retail sales tax offers complexities and problems for both the administrative agencies and the taxpayers. These problems cannot be overcome within a few months, and a retail sales tax should not be adopted with the thought that it could be dropped in six months or a year, if it is hoped that its administration will be effective. For its successful enforcement and equitable application, the tax should be planned carefully, a competent and adequate administrative force should be developed, and a few years of experience should be allowed for the smooth operation of the law. The tax could not be collected successfully without the hearty co-operation of the vendors and the consumers, and this co-operation could hardly be won overnight.

THE ROLE OF THE SALES TAX

I have always looked upon the sales tax as an emergency fiscal device which should be employed only after the traditional income tax and the excises have been developed to their practicable limits. The tax suffers from inequalities in its burdens whether it is shifted to consumers or absorbed by sellers, it is politically unpopular with organized labor and with other groups and persons, its administration is not easy, its revenue possibilities are limited, and it would have some inflationary effects if it were adopted now or during the post-war boom.

On the other hand, many inequalities could be avoided by exempting food and by not applying the tax to rentals and payments for homes or to essential medical and other services, and its regressive burdens could largely be counterbalanced by the use of progressive taxes. The tax would aid in a minor way in lessening the pressure of inflationary consumer spending, it would be popular with many persons in the middle and upper income brackets and with some in the lower incomes, and its administration should be practicable.

The choice here of tax measures to curb inflation and raise revenue in wartime must be made among the available alternatives, which include the extension of the personal income tax, more reliance upon the excises, the introduction of a spendings tax, the enactment of a sales tax, or some combination of these measures. The advantages and the disadvantages of the alternatives should be studied, and the choice should rest with the tax or combination of taxes which are most advantageous, considering not only the taxpayers but the entire community.

In peacetime, the role of the sales tax is more uncertain. Its social effects during depressions are undesirable. If the tax is employed, its use should be limited to periods of prosperity, if possible.

FISCAL POLICY AND SOCIAL REFORM AS IT MAY AFFECT THE POTENTIALITIES OF THE PERSONAL INCOME TAX

HENRY SIMON BLOCH[1]

University of Chicago

(*Read February 18, 1944, in the Symposium on Taxation and the Social Structure*)

THIS paper is divided into two major parts. The first part is a discussion of certain phases of the present state of the debate in the field of fiscal policy and the social outlook of modern fiscal analysts. It is entitled "The State of the Debate." The second part, entitled "The Potentialities of the Personal Income Tax," is concerned with the more pragmatic problem of combining the three major fiscal reform proposals of our time into one single instrument. Forced savings (Keynes plan), the social security plan (as outlined with some variations in the Beveridge plan and in the Wagner-Dingell Bill), and the generally emerging notion of the "ideal" personal income tax can be merged into one fiscal program. If the social security system is made universal, *i. e.*, if everybody can be included, there is no reason for the separate existence of a payroll tax and an income tax, particularly since the latter is also based on a witholding plan. This would also be the first step towards financing social security out of general revenue and towards abolishing the "assignment of revenue" system. Forced savings may well take on the form of special old-age benefits, unemployment insurance, and medical insurance (preferably beyond certain minima which the government may guarantee to everybody regardless of taxes paid in the past). The personal income tax, if simplified and revised, could well become the nucleus of the fiscal revolution of our day.

THE STATE OF THE DEBATE

While most of the modern Anglo-Saxon authors in the field of government finance refrain from rigorous analysis of social philosophy,[2] they consider fiscal policy from a broader viewpoint than that of the Latin or German economists of a few decades ago.[3] Hardly any author of consequence today would think of dissociating the effects of taxation from those of borrowing and spending or of ignoring the effects of taxation on the national income.

It is quite obvious that while the economic analysis of our present-day Anglo-Saxon authors is more rigorous, the social and political philosophy of the analysts is generally vague. To a superficial observer of the Anglo-Saxon and Scandinavian literature of the thirties it must appear that cyclical fiscal policy has become a common bond linking what can generally be called the Keynesian - New Deal - Social Democratic group. Cyclical fiscal policy became in the thirties as much of a battle cry as collective bargaining and free competition. It became a fighting word like "progressive taxation" was about half a century ago. Still, cyclical fiscal policy is not even part of a social philosophy. It is a fairly well defined means leading to an end which lacks any definite sort of description. If its purpose is to smooth out the cycle, is it also to provoke any lasting change in the economic system? Must it change the body economic? Must it precipitate social changes? Will it slowly change the status quo?

The pale *Economic Program for American*

[1] The author thanks Professors L. W. Mints and H. C. Simons for some good suggestions.

[2] This does not mean that modern writers do not recognize the need for a social philosophy. *Cf.* Simeon E. Leland, "The Theory and Practice of Taxation," *Proc. Nat. Tax Assoc.* 1939; 696–709; or *cf.* in the present publication the article by Roy Blough.

[3] Benvenuti Griziotti, Italian student of government finance, held that the study of government finance should only be concerned with revenues, since public expenditures were not determined by strictly economic considerations. *Cf.* his *Considerazioni sui metodi, limiti e problemi della scienza pure delle finanza* (Rome, 1912). This and similar views by other authors are presented by Gaston Jèze in *Cours de Finances Publiques*, 1936–1937 (Paris, 1937): 4–7. Jèze himself emphasized that government finance should be considered broadly, including loans and taxes, spending and budgeting. It should take into account factors which are political, social, legal, economic in nature. *Cf. ibid.*: 7–19. The reader of Jèze's writings, however, will notice that the effects of borrowing on national income are nowhere properly analyzed, that the interrelationship between taxes and prices is treated incompletely, that the connection between monetary theory and fiscal policy is not recognized.

Democracy[4] and pamphlets like *After Defense What?*[5] were too fragmentary to indicate the new trend. Lord Keynes's *General Theory of Employment, Interest and Money* (1935) was too far removed from political analysis. Political writings on taxation, such as Sidney Ratner's *American Taxation; Its History as a Social Force in Democracy*,[6] did not take into account the new developments in the field of economic analysis. Alvin Hansen, after a series of rather timid writings, came out with *Fiscal Policy and Business Cyeles*,[7] in which he tried to fit the new trend in fiscal policy into a structural analysis of the American economy.

Henry C. Simons, who came to "bury Hansen,"[8] recognized the importance of the book.[9] The essential controversy is political rather than economic, although there is much terminological shadow-fighting on both sides of the fence. Anybody who reads Simons' article will notice how he voices his anger in view of the fact that Hansen makes too many compromises, that he gives him so much trouble in classifying friend and foe.[10] Yet Simons, who calls himself an "unreconstructed liberal," sees the real difference, which is political in nature. The conflict between Hansen and Simons, as expressed in Simons' biting review article, is indicative of the type of the present controversy.

Essentially there are some important economic points on which Simons disagrees with Hansen. The differences come out sharper if one includes Simons' "Rules versus Authorities in Monetary Policy"[11] in the comparison. Hansen and Simons are both for monetary expansion in periods of depression.

It would be wrong to believe that laissez-faire economist Simons would underestimate the need for monetary expansion in a period of depression. In fact, he goes much farther than Hansen. Where Hansen advocates only mild deficit financing, Simons prefers straight issuing of currency.[12] Simons does not deny the value of the economic effects of note issue on purchasing power during depressions, yet he fears that a steadily increasing national debt would eventually lead to some kind of a revolution, if not by the weary taxpayers, at least by the rentiers who are supposed to get the jitters sometime in the dim future.[13] It is a political fear—revolution—rather than an economic argument which motivates the objection. The political conflict between Simons and Hansen on this particular issue is not typical of the conservative-progressive or the laissez-faire - collectivist controversy. Most "progressives," even "collectivists," would prefer printing money to borrowing from banks and wealthy individuals.[14]

H. G. Moulton recently published a much discussed little book, entitled *The New Philosophy of the Public Debt*.[15] This publication is not directly comparable to Simons' article. Catering to a different public and written in popular fashion, it can be mentioned as one of the most talked-about contributions to the anti-Keynesian literature. The book was well reviewed by Professor David McCord Wright.[16] Obviously and unfortunately Moulton appears as a rabid and uncompromising budget-balancer. Careful reading of the book, however, reveals, as McCord Wright put it, "that Moulton is in far greater agreement with his opponents (for example Professor Hansen) than he ever admits."[17] Whatever agreement there is between Moulton and Hansen is confined to matters purely economic. Whatever disagreement there is, is primarily determined by political differences even when such differences are stated in economic terms. This may seem an oversimplification, but grains of salt can be liberally added without changing the taste of the dish.

In the late thirties, it seems as if the Keynesian

[4] *An Economic Program for American Democracy by Seven Harvard and Tufts Economists* (New York, 1938).

[5] See also other National Resources Planning Board publications.

[6] New York, 1942.

[7] Alvin H. Hansen, *Fiscal Policy and Business Cycles* (New York, 1941).

[8] "But I come to bury Hansen—albeit respectfully and despairingly" (H. C. Simons, "Hansen on Fiscal Policy," *Jour. Polit. Econ.* **50** (2): 162, 1942).

[9] "Few books in our literature can offer so much in terms of useful information and ideas or do more to bring readers abreast of important recent inquiry. . . . It may well become the bible for that substantial company of intellectuals, following Keynes and recklessly collectivist, whose influence grows no less rapidly in academic circles than in Washington." (*Ibid.*)

[10] "Indeed, it is admirably contrived to intensify already bitter controversy, to confuse what should be a sharper division along individualist collectivist lines, and to divide us as supporters or opponents of a particular administration" (*Ibid.*: 184).

[11] *Jour. Polit. Econ.* **44**: 1–30, 1936.

[12] That comes out clearly in Simons, *ibid.*

[13] "The fright of bondholders will create a revolutionary situation" (Simons, *Jour. Polit. Econ.* **44**: 175, 1936).

[14] A. P. Lerner, whom Simons would certainly call "collectivist," advocates printing money as an expansionary measure. (*Cf.* A. P. Lerner, "Functional Finance and the Federal Debt," *Social Research* **10** (1): 41, 1943.)

[15] Washington, D. C., 1943.

[16] *Am. Econ. Rev.* **33** (3): 573–590, 1943.

[17] *Ibid.*: 573.

theory had been adopted by the Administration. It does not follow that the New Deal philosophy, any philosophy, has been adopted by the Keynesians—all the Keynesians. Nor does it follow that the Administration, when hamstrung by a recalcitrant Congress and facing the task of fighting inflation, followed the Keynesian advice in matters of taxation. Nevertheless, the economic controversy over Keynes coincides today with a political difference of opinion, however vaguely defined.

Conservatives attacked Keynes and his followers because Keynes attempted to change the status quo of the distribution of wealth and also because he threw their sacred rules of "balancing the budget" into the gutter of vulgar fallacies.

Progressives followed Keynes because his recommendations tended to increase the standard of living and because they tended to make orthodox economists the laughing-stock of the profession. At times and in many instances progressives opposed Keynes, particularly so Labor Unions when Keynes came out for a stiff program of war taxation combined with deferred pay. Yet the progressive opponents of Keynes were in the political camp, while the progressives who belonged to the economic profession and particularly the "fiscal analysts" formed a rather solid block behind him. They were for credit expansion during the depression, they are now for taxation and non-inflationary borrowing during the inflationary period.

It is undoubtedly true that all those who rally to the banner of the Keynesian school have common ideals and subscribe to the Keynesian theory because it indicates some of the ways and means to achieve a permanent stage of full employment, a fuller use of resources, an avoidance of ever-recurring deep depressions. It also implies a steady shifting in the distribution of wealth from the top to the bottom classes. Keynes's war finance plan really implies a reversal of the old method of "taxing the poor and borrowing from the rich" by inaugurating a scheme of "borrowing from the poor and taxing the rich." In its final implications the Keynes plan includes also such socialist hobbies as a capital levy and a total redistribution of wealth.[17a]

It is obvious, however, that one can still be a good "Keynesian" in the commonly accepted sense without supporting anything of Keynes's recommendations besides equilibrium analysis and cyclical fiscal policy. The method of equilibrium analysis as such is unrelated to any political credo. It is a method of analysis rather than a recommendation of policy. Cyclical fiscal policy as propagated by Keynes leaves enough room for the most different tastes.[18] In fact, it is only valid for a capitalist system where the business cycle with all its phenomena plays its game. It is designed to eliminate this cycle and incidentally to achieve a redistribution of wealth.

It is not possible to imply that any type of cyclical fiscal policy will lead to a change of the economic system. The cycle itself is too vague a phenomenon to be taken as an object of a consistent policy. Even if it were much more clearly recognized, it still would not be sufficient to conquer the cycle. It is not possible to say that "cyclical fiscal policy will lead to another system." It is only possible to say that a certain type of cyclical fiscal policy may lead to or precipitate a development toward a change of the status quo. The fact is, that if the depression is not tackled at all, a change of the status quo will be achieved without any government action.

It would not be futile, however, to ask the outstanding advocates of "cyclical fiscal policy" in which direction they wish to influence the social change. While Hansen implies that cyclical fiscal policy is designed to maintain free enterprise, he does not clarify what he means by free enterprise. His conception may be very different from the conception of Henry Simons. Yet Simons is quite clear in what he wants, while the chief advocates of cyclical fiscal policy in the Keynesian sense are remarkably evasive in re institutional forms.

There is no doubt that the influence of Keynes and his followers (taken in a very wide sense) has contributed to a better understanding of economics by government officials. In general (not considering the tendency of the individual advisers) economists who acted as government advisers not only sold some of their pet ideas but have taught leaders of the government to talk in terms of economic analysis. During the last war many economists knew that taxation could be used for inflation control. Yet Secretary McAdoo consistently emphasized the revenue as-

[17a] *Cf.* John Maynard Keynes, *How to Pay for the War: A Radical Plan for the Chancellor of the Exchequer* (New York and London, 1940).

[18] Theoretically the recommendations for cyclical fiscal policy are a result of conclusions based on equilibrium analysis. Very often, however, superficial Keynesians have accepted cyclical fiscal policy without even knowing about equilibrium analysis.

pect of taxation alone. Today, it is a matter of course that the President, the Secretary of the Treasury, the Director of the Budget, and other officials should talk in terms of fiscal control.

Arguments like "what you pay in higher taxes, you will not pay in higher prices" are being used in government declarations. The President's Budget Message of 1941 is an example of fiscal common sense. Obviously, those who administer the country's finances have learned economics.[19] And a good deal they learned from Keynes, directly or indirectly.

The question is whether the Keynesian scheme is socialist by definition because it implies a redistribution of wealth. Of course it reminds one of the theory of Adolf Wagner with the essential difference that Keynes has a wider concept of fiscal policy than Wagner had, because Keynes recognizes the effects of borrowing as a factor in redistribution and he considered the effects of fiscal policy on the general equilibrium situation.

The essential difference between Keynes and "tax socialists" is that Wagner and Hobson advocated only scaling down from the top, while Keynes advocated lifting up from the bottom as well. Both attribute to fiscal policy a key role in the social reform of their times.

It must be remembered that Marx and the original Marxists gave only little importance to the topic of taxation. Karl Marx, of course, minimized the importance of distribution. A few quotations will recall his attitude.

> It is wrong anyhow to make much of a case of the so-called distribution. . . . Every distribution of the means of consumption is only a consequence of the distribution of the conditions of production itself. This latter consequence is, however, a character of the production system itself.[20]

More specifically on taxation, Marx wrote:

> The relationships of distribution which are based on the bourgeois (system of) production, the relationship between wages and profit, profit and interest, land rent, and profit can but be modified by the tax in their unessential points, but never can they be menaced in their basis. Even the abolition of taxation could only further the development of bourgeois property and its contradictions.[21]

No matter how favorably one is inclined towards the Marxian writings, it remains a fact that Marx did not show greater wisdom in matters fiscal than the other economic writers of the time.

Fiscal policy was not part of the stock in trade of the scientific socialists. It became, however, a major element of the intellectual ammunition of latter-day reform socialists. With much improvement in fiscal policy, with infinite improvement in economic wisdom, with little added to political philosophy, the Keynesians took over much of the heritage of the "tax socialists" of the last century. Not that there is much of a *filiation d'idées;* it is rather a development which takes place without much conscious contribution to the process by the present-day representatives.

THE POTENTIALITIES OF THE INCOME TAX

In their practical programs, Marx and Engels, unlike the utopian socialists,[22] advocated heavier taxes. A heavy progressive income tax was proposed in the *Communist Manifesto*, and a capital tax was recommended by Marx in journalistic articles.[23]

It was at the Congress of Eisenach (1869) that the German socialists proposed to "abolish all indirect taxes and introduce one single direct and progressive income and inheritance tax."[24]

This sounds strangely like one of the writings of my anti-socialist colleague, Henry C. Simons. He too advocates that the personal income tax be the mainstay of the tax system and that inheritance and gifts should be treated as income.[25] The resemblance is of course "purely coincidental and accidental." The socialists of Eisenach did

---

[19] It must be noted, however, that the bond campaign is frequently carried out with wrong arguments. It is often stated that "bonds will shorten the war." It is rarely stated that bonds will help in the fight against inflation. The wrong arguments are mostly brought forth by the Hollywood group, seldom by official representatives of the Treasury.

[20] Karl Marx, *Neue Zeit* **9** (1): 568. Translated from Fritz Karl Mann, *Steuerpolitische Ideale* (Jena, 1937): 285.

[21] Translated from Karl Marx, review of Emile de Gerardin, *Le Socialisme el l'Impôt* (Paris, 1850), in F. Engels und F. Lassalle (ed. F. Mehring), *Aus dem literarischen Nachlass von Karl Marx*, ed. 2, **3** (Stuttgart, 1913): 435.

[22] Thomas Morus, Campanella, Louis Blanc, and others advocated the tax-free state. Babeuf advocated the freedom from taxation for citizens only, recommending that tributes be levied from subjected enemies. For a discussion of the Utopians, read F. K. Mann, *Steuerpolitische Ideale* (Jena, 1937): 280–284.

[23] It must be emphasized that tax proposals meant only an immediate minimum requirement of the German socialists. For the long run they demanded more drastic measures.

[24] Point 9 of the "Eisenacher Programm" (*Cf.* Mann, *op. cit.*: 303).

[25] Henry C. Simons, *Personal Income Taxation* (Chicago, 1938): 211.

not share many of the views which Simons presents today.

In the tax debate of today the barriers between the "left" and the "right" (excepting the extreme left and the extreme right) tend to disappear among professional economists in this country.

Some of the most vigorous advocates of deficit spending have turned into very emphatic advocates of increased taxation. They believe in the use of different financial means at different times.[26] Among the advocates of anti-inflationary taxation, one finds the overwhelming majority of professional economists, economists of all shades; "rightists" and "leftists" and the middle-of-the-roaders agree on this technical necessity. Those who oppose a tax program of considerable magnitude usually have an axe to grind. The President, Mr. Morgenthau, Randolph Paul, and others were outspoken enough in their reflections upon the reaction of the Ways and Means Committee.[27] The strange element in the matter is that the most vigorous congressional opponents of public borrowing do not advocate higher taxation, which is the chief alternative to increase in the "size of the debt." The reaction of many Congressmen, journalists, and laymen notwithstanding, professional economists agree on the main that more and bigger taxes are needed. They disagree, however, as to what type of taxes should be applied. The most clear-cut statement of disagreements can be found in *When Authorities Disagree—Financing World War II*, by Simeon E. Leland and Harley L. Lutz.[28] Among professional economists Harley L. Lutz is one of the few advocates of a gross income tax as one of the chief means of financing the war. Leland comes out for heavier reliance on personal income taxation and deferred pay, but is not opposed to a sales tax in addition to all that.

There is an overwhelming majority among professional economists who favor some type of graduated personal income tax as an important part of the tax structure of the country. In 1936 a poll was made by Mr. Mark Graves, then President of the Tax Commission of the State of New York. A number of university professors and "authorities from other walks of life" were polled as to whether they favored certain taxes. The largest number of unqualified "yes" was given with regard to the graduated personal income tax. There were 126 "yes," not a single "no," and only one reply was so full of doubts and qualifications that the compilers marked it as uncertain or undecided.[29]

There is hardly anybody among professional economists who would demand the elimination of the graduated personal income tax. Most economists, however, want to modify the present system of income taxation.

The more conservative elements among economists, consultants to "economy leagues," trade associations, large industries, want a sales tax in lieu of a stronger income tax. The more liberal or rather the independent elements want all kinds of taxes, but every one of them wants a considerable increase in the use of the personal income tax.

It will not be necessary here to go into the differences between sales tax and income tax which have been repeated over and over again. Nor will it be necessary to explain that a sales tax would be better than no tax increase at all, because the alternative would be a further rise in prices. It is, however, important to get to the fundamental question as to who should be protected in a period of inflation.

Obviously, inflation hits everybody who cannot increase his income during a period of rising prices sufficiently to maintain his standard of living. Strangely enough, the people have swallowed the faulty generalization, apparently following from the above argument, that the middle class is hit most. This generalization is made because fixed-income groups cannot increase their income, and since most salary-earners are in the so-called middle-class group, the middle-class people seem to be the chief sufferers. An additional false conclusion points in the same direction: Inflation wipes out savings, middle-class people have savings, therefore they suffer most.

It is important to remember that inflation, a rise in prices, means a reduction in real income for those people who cannot get sufficient increases in money income. Losses in real income are more serious for poor people than they are for middle-class people. The poorer an individual, the more serious are the sacrifices he has to undergo if he receives further cuts in income. The lowest-income class may have to deteriorate their diet, the second lowest their clothing, the third

[26] The most outspoken formulation of this principle is found in A. P. Lerner, "Functional Finance and the Federal Debt," *Social Research* **10** (1), 1943.

[27] The veto of the tax bill came after this paper had been delivered.

[28] New York, 1942.

[29] *Cf. Tax Systems of the World*, ed. 7: 412, 1938. The only answer classified as undecided was that of Professor F. R. Fairchild, of Yale University.

lowest their shelter, and the fourth lowest perhaps only their amusement or "prestige expenditures." If there is anything like competition among buyers, the poor buyers must be protected against the middle-class buyers, and the middle-class buyers must be protected against the rich buyers. If this is taken for granted, one must have very clearly in mind that in America there is a high social and economic mobility and that not all of the poor of yesterday are the poor of today. It is the poor of today, no matter what their profession or trade is, who must be protected.

A stiff sales tax without exemptions might well kill the patient. A sales tax with exemptions might bring some relief, since it is better than no tax at all.[30] Since, however, the essential difference between an increase in taxes and an increase in prices is, that in the first case one can plan the distribution, while in the second distribution will be unplanned, the argument can be clearly made from an objective point of view. If one is committed to a control of the socially inequitable inflation, and if one is furthermore committed to a distribution control which provides protection of the economically weaker against the economically stronger, the graduated personal income tax is appropriate because it takes higher percentages out of larger incomes. To those who object to any kind of redistribution, it may be told that whenever total national output of consumer goods for the people is fixed, changes will occur. Either the rich get richer and the poor become poorer, or the poor get more and the rich get less. I, personally, advocate the latter as the only honest democratic solution for the distribution of sacrifices. It will not wipe out all inequality, but then, the author of this paper is not committed to an equalitarian system. It will, however, provide the more democratic of a number of less democratic alternatives. The sales tax, no matter how much improved, takes higher percentages from low incomes than it takes from high incomes.

The income tax, as it stands, needs a great number of improvements. First it needs simplification and streamlining. Professor Leland put it this way in *The United States News*.[31]

> The most important single simplification in the federal income tax law would be the adoption of one scale of progressive rates. By incorporating the normal, surtax and Victory tax rates into a single rate schedule, one computation would replace the three now required. The real difficulty about integrating the present rate schedules is in taxing the income from certain outstanding federal securities, some of which is subject only to surtax rates.
>
> Again, the average taxpayer must suffer for the preservation of an unwarranted tax immunity. Perhaps the surtaxes could be preserved only to reach income from tax exempts, everyone not so affected being given the advantage of a simple rate progression. The next step would be to apply the rates to taxable specific sums, appropriate to the circumstances, from the liabilities so computed, rather than by deducting fixed amounts from taxable income.[32]

In addition a number of other reforms are necessary. Separate returns, particularly in community property states, leave room for special privileges and should be replaced by compulsory joint returns. Automatic percentage deduction for oil wells and many mining industries is unjust. Owners of real estate pay no tax for income received in kind, since they enjoy tax-free rentals in houses built with their own capital. Income from inheritance and gifts should be taxed like income from work or investments.

If an increase in the rates of the personal income tax is adopted, and if, say, the lowest group, from 1,200 to 2,000, will have to pay 30 percent as income tax, many people will suffer real hardship. It is not a hardship that requires special consideration if a man owns municipal bonds. It does need special consideration, however, if a man is sick or if a child is born or if a death occurs in a family. Future privileges should be bestowed on the basis of hardship rather than on the basis of economic feudalism. The people should be informed that in case of sickness they may demand tax-waivers on a much more generous basis than they receive at present; that on the other hand unwarranted loopholes in the tax system should be stuffed. The higher the rates, the more serious becomes the problem of "hardship waivers." The problem of hardship waivers is closely connected with the problem of family budget planning. Medical expenses, temporary unemployment, death and birth can upset the most careful planned budget.

---

[30] On minimum exemption schemes, *cf.* Henry S. Bloch, "For a Just War Finance Policy," in M. Ezekiel, M. Bondfield, A. Meiklejohn, Caroline Ware, et al., *Toward Democratic Victory* (Washington, 1943): 9.

[31] February 25, 1944, p. 36. "To present a cross-section of authoritative opinion, the United States News asked Congressmen, economists, lawyers and others this question: What is the most important change in the individual income taxes that Congress could make to simplify those taxes?"

[32] *Ibid.*: 36, 38.

There is no need here to dwell on the merits of social security. Suffice it to say that the author is committed to the principles of the Wagner-Dingell bill. The author believes that now is the time to start with a large-scale social security program. Since it is undoubtedly necessary to start as quickly as possible with the enactment of higher tax rates on the lower as well as on the higher tax brackets, the question arises as to whether or not the entire savings of the lower-income brackets should be taken away by the government. The alternative is, of course, to have forced savings or deferred pay for the lower-income groups. Forced savings instead of higher taxes, higher prices, or both, should not be too unpalatable an idea for the great masses of the population. One of the chief problems in connection with forced savings arises in the matter of redemption. Redemption must, naturally, be at the discretion of the government, and it should be timed according to the requirements of the cycle.

Social-security deductions and forced savings have two things in common. They are both present payments for future benefits. To cope with the inflation of today we must reduce consumer spending. While forced savings may not provoke a reduction of consumer spending corresponding dollar for dollar to a possible reduction resulting from taxation, they nevertheless will closely approximate such a reduction, particularly in the lower brackets.[33] Increased social-security taxation will most certainly produce a reduction in consumer spending. Increased benefits will not materially change the general economic situation. The disbursements will not be very large for some time to come, because the "subscribers" will need some time to accumulate substantial claims. Claims will become important when the first wave of unemployment will hit the country. This may be a period when an increase in consumer spending is desirable. Yet, even if it is not desirable (because of production difficulties due to reconversion), relief payments will be appropriate and necessary. (Any inflationary effects can be outbalanced by a courageous tax program affecting the income-earning elements of the population, coupled with continued price control and rationing.)

It is recommended by this author that the social-security system be made universal. It is furthermore recommended that it be incorporated into the personal income tax instead of the social-security tax system, particularly considering the fact that our withholding system is being well developed.[34]

Such a move would couple a number of important fiscal reforms. It would increase the taxes in the lower brackets, while offering a *quid pro quo* to the payers. It would lay the foundation of a post-war social-security plan of considerable magnitude.[35] It would make it possible to use the personal income tax as an instrument of fiscal and therefore social policy of great effectiveness.

[33] The method of forced savings does not produce necessarily anti-inflationary effects when applied to middle or higher income brackets. In income brackets where voluntary saving is ordinarily practiced, forced savings would only replace other savings. In lower income brackets where people ordinarily do not save, forced savings cut down consumption. Forced savings are a sound alternative to taxation only where low income brackets are concerned.

[34] The ideal system would be to include social-security financing in general expenditures. The author does not wish to further the idea of assignment of revenues. He only recommends an alternative to financing social security from payroll taxes by financing it from income taxes. There should be a possibility for taxpayers to accumulate limited benefits over and above a minimum that the government should grant to everybody regardless of contributions. To put it more precisely: the social-security plan should not be dependent upon income-tax yield. Income-tax payments should, however, be partly transformed into contributions for social insurance. A high percentage of the payments in the lower brackets and a small percentage of the payments in the upper brackets might be thus earmarked.

[35] Servicemen should, of course, be included in the plan and should receive special consideration.

THE HARMONIZATION OF FISCAL AND SOCIAL AIMS

MERLIN H. HUNTER

University of Illinois

(Read February 18, 1944, in the Symposium on Taxation and the Social Structure)

THE motive assigned for the levy and collection of taxes varies from one individual to another and from group to group. Many fiscal officials emphasize that their sole concern in the levy of taxes is to raise revenue, while the social and economic reformer would use the device of taxation as a means to effect changes previously determined to be desirable for the welfare of the group. In addition, of course, there are those who see in taxation a device through which they may accomplish their own selfish ends. With this latter motive we need not be concerned except to condemn it, but the two others are so intertwined and frequently so obviously antagonistic that their relationship is deserving of serious consideration. The administrator who contends that his only concern with taxation should be with the receipt of revenue, while he will leave it to the philosopher, the economist, or the sociologist to give attention to the social consequences, has been very superficial in his thinking. Whatever the statements, the fact is that every tax, notwithstanding the motive for the levy, will have social and economic repercussions. Since, then, the revenue and social aspects of taxation cannot be divorced, harmony rather than friction in their relationship is highly desirable.

No one should have the impression that interest in the social consequences of taxation is of recent origin nor that, in placing renewed emphasis upon this aspect of taxation, many new principles have been evolved. It may be helpful to review a few ideas of previous centuries, yet which seem to possess an extremely modern flavor. Many biblical references, it need scarcely be noted, are made to tolls and tributes, as well as to methods of levy and exemptions. Concern for justice is exemplified in the "Sacred Laws of the Aryas" in that those who were to be exempt from taxes were listed while the rate upon those subject to taxation was made to vary. In the Laws of Manu a warning is found the importance of which has not lessened with time: "Let him (the King) not cut his own root nor the root of other (men) by excessive greed." In these Laws, also, we get an idea of maximum taxation in: "A King who in times of distress takes even the fourth part (of crops) is free from guilt if he protects his subjects to the best of his ability."

In the half century before Christ the Romans and Greeks were concerned with justice in taxation. According to Herodotus, officers were to be appointed to insure that no man might be taxed beyond his ability. Plato concluded that the public officers might take a certain portion of the whole value or of the annual revenue, whichever they deemed best. According to Livy, public expenses were to fall upon the rich, since the education of their children would be a sufficient tax upon the poor. Dionysius of Halicarnassus and Solon, moreover, may have been the instigators of an argument, an agreement to which, acceptable to everyone, has not yet been reached. The former declared for proportionate taxation, the latter for progressive rates.

As centuries passed, taxation occupied a place of growing importance in the thinking of those who were concerned with human relationships. Hobbs, for example, contended that taxes, in general, should be paid on the basis of benefits received. This led him to conclude that equality of burden would be attained when people paid taxes according to their expenditures. Harrington in the seventeenth century saw in taxation a means to accomplish other than revenue, as is indicated by a statement from his Oceana:

> But if a man has ten children living he shall pay no taxes; if he has five living he shall pay but half taxes; if he has been married three years or be above twenty-five years of age, and has no child or children lawfully begotten, he shall pay double taxes.

Locke was of the opinion that taxes, however contrived, for the most part terminate upon land. Hume would levy taxes upon consumptions, especially those of luxury, because such taxes are least felt by the people. Montesquiev would levy taxes upon goods because the people would hardly know they are paying. Physical neces-

sities, he thought, should not be taxed; useful things should be taxed, but less than superfluities.[1]

A great deal of space could be devoted to a portrayal of the early ideas on taxation, but a more pertinent and profitable consideration lies before us. We have sought merely to demonstrate (1) that, from the interests of many of the early philosophers, a discussion of taxation should not be out of place before the members of this Society, and (2) that a modern student of problems of taxation need not hope to do much in the development of new concepts concerning the methods of taxation; the most he can hope to do is to apply the old concepts to a more complex social and economic order than that which existed centuries ago.

It would be futile to attempt to estimate the amount of the post-war tax levy, but we may be sure that any previous peacetime levy will be insignificant in comparison to it. The interest charge upon the federal indebtedness alone will be greater than the total federal expenditure in 1930. At present we cannot predict the extent to which the activities of our economy will be left to the direction of private initiative or to what extent they will be dominated by the dictates of government. It is certain, however, that we will experience neither "free enterprise" with no restrictions from government nor an economy void of any frontiers and therefore in such a state of maturity that its operation should be planned by the government.

Our economy has become too complex either to "take the government out of business" or to return to "free enterprise." Certainly governmental units will continue to provide goods and services which could be secured through private enterprise; just as certainly private enterprise will function under a certain amount of regulation. We should insist, however, that the regulation be helpful to the conduct of business rather than destructive of it. In providing rules and umpires for athletic contests, the responsibility is placed upon those who have been participants in the games and therefore are familiar with the problems which are encountered. Regulation of business might be more constructive if more attention were given to the same policy.

If, because of a mature economy, it would seem desirable for the government to "take over" and provide "economic planning," "full employment," "social security," "social planning," and the many other functions implied in slogans which have been heralded and sung in recent years, then our whole social and economic system will be harmonized into a millennium through all-out taxation and the problem implied in our topic will have been solved. Few of those who have given considerable thought to the future, however, are willing to admit that our economy has reached such an impasse that any such drastic change is indicated. Geographical frontiers of the future may indeed be limited, but no evidence exists to indicate other than that our mental frontiers lie far beyond the horizon. A few suggestions have come to us of the future possibilities in the fields of electronics, radar, aviation, and plastics; indeed, if all mental explorations into new frontiers within the last few years which are as yet "military secrets" could be portrayed before us, they would act as a "block-buster" upon the claim of a mature economy and with it the extinction of the need for private initiative.

We may safely assume, then, that in our post-war economy many functions will be provided by government for which the levy of heavy taxes will be necessary; that private enterprise will continue to function and upon this taxes will be levied. Doubtless, moreover, taxation will take a larger part of the total and per capita social income than in any other peace years in our history. This but emphasizes the need for a closer scrutiny of the relationship between the fiscal and social aims of taxation.

Before the utmost can be attained in a harmonious relationship, the individual must be willing to accept the philosophy that the good of society must supersede the desires of individuals, or the groups of individuals, which compose it. That this is a well-defined American principle needs no extended proof. We have witnessed its operation in municipal zoning ordinances, in the regulation of traffic, in the location of slaughterhouses and liquor establishments, and in the protection of women and children in industry. In the levy of taxes, however, legislators, often aided and abetted by court decisions, have frequently grievously transcended this principle. Tax legislation has not been free from the influence of the familiar pressure group.

Taxes upon oleomargarine, filled cheese, and mixed flour may be taken as examples of the levy of federal excise taxes primarily to effect changes for the benefit of individuals to the

[1] For an extended discussion of the development of the ideas concerning taxation, see Robert Jones, *The Nature and First Principles of Taxation*, ch. 2, 1914.

detriment of society. In the arguments before committees and in the debates in Congress the revenue aspects received practically no attention, but the Supreme Court took the position (*McGray* v. *United States*, 195 U.S. 27) that the levy was properly an excise and, furthermore, the "motive or purpose of Congress in adopting the acts in question may not be inquired into." In a later case arising out of a tax levied upon goods produced under conditions of child labor (*Bailey* v. *Drexel Furniture Company*, 259 U.S. 20), the Court took the position that the Act was more than a tax, that it was designed as a penalty for certain business practices and therefore unconstitutional. Here the Court inquired into the motive back of the tax. In its decisions in these, as well as in many other tax cases, the Court has not seemed to be greatly impressed by the jeweled nature of consistency. In the levy of taxes upon oleomargarine and chain stores, as well as in other cases, the states have prostituted the device of taxation to direct the economy for the benefit of particular individuals. In general, no tax can be justified the purpose of which is to effect or to prevent an economic or social change which will benefit particular individuals at the expense of the social group. As long as sanction is given to such a practice, a harmonious relationship between the fiscal and social aims of taxation cannot be attained.

In the solution of our problem we must assume the role of a pioneer who seeks a tax system which will produce sufficient revenue for the justifiable functions of government and yet which will mesh with as little friction as possible into our social and economic structure. In this we may start our quest with the enthusiasm of Sir Launfal in his search for the Holy Grail or of Ponce de Leon in his attempt to discover the fountain of youth. Others, of course, have gone on the same quest, and few are unfamiliar with the perfect tax system proposed by Henry George and another proposed by Doctor Townsend. The adoption of either of these, according to the claims of the "discoverer," would result in a social and economic millennium; in other words, we would achieve complete harmony. Those of us, however, who are not convinced of the validity of these discoveries must seek further.

Lest I be accused by the more practical-minded of being merely an "armchair philosopher," let me hasten to assure you that I am aware that, in any proposed tax, due consideration must be given to administrative feasibility and, I hesitate to admit it, political expediency. Neither of these, however, should receive consideration in the first instance. Attention should first be directed to the fairness of the tax and to the social and economic repercussions. Administrative feasibility and political expediency should only modify these basic considerations but should never replace them.

The first procedure in our quest for a harmonious tax structure must be to retrace some of the steps that we have taken. Too frequently a tax has been introduced because of the exigencies of the Treasury while little thought was given to social or economic consequences. As years have gone by, additions and modifications have been made, often magnifying the undesirable features of the first enactment. Then, too, a changing economic and social structure may demand a change in the system of taxation which has not been accomplished. The constitutions of many of the states contain provisions which may have been pertinent to an economy of a century ago, but when an attempt is made to apply them to the complex social and economic structure of today, confusion, injustice, and dissatisfaction are the outcome.

The present method by which the federal government taxes the income both of individuals and of corporations illustrates the degree to which injustice and friction may grow in a comparatively short period of time. The tax upon personal incomes as levied at present violates two of the cardinal principles set forth by Adam Smith, certainty and convenience. In this emergency, all the evidence points to the fact that the people do not object to the demand made upon them to pay taxes, but that they do object strenuously to the complexity, inconvenience, and uncertainty of the present method of levy. Granted that some of the regulations are aimed to secure more minute justice; yet any satisfaction from the attainment of justice has been more than offset by dissatisfaction and friction, indicating that the regulations have been ill-advised. Granted that many of the provisions are designed to bring more dollars into the Treasury; yet if this result is more than offset by delays and increased costs of administration because of the complexities, then such provisions cannot be defended. The growing emphasis upon the slogan, "simplify the income tax," may bring results. The taxpayer is probably more interested in simplicity and certainty than in the few dollars' difference in his tax which the present system may mean. It is

possible greatly to simplify the determination of net income, while a single scale of progressive rates could be made to apply to the total net taxable income in different predetermined brackets.

The present indefensible treatment of corporations from the standpoint of taxation illustrates the condition which may arise when taxes are imposed with no consideration of the merit which may be involved. In the first place, the corporate form of business organization is no more inherently good or bad than is that of the partnership or individual enterpriser; consequently regulation could not be assigned as the motive for taxing this particular form of business organization. The existence of shares of stock made it feasible to administer a corporate tax, but once emphasis is placed upon earnings as a base for taxation, this reason is no longer valid, since forms of business other than the corporation also have earnings. The reason cannot be that the earnings of corporations are always larger than those for other types of business, for the facts do not warrant such a conclusion; if this were true, moreover, a tax upon the earnings of business of whatever form of organization would fall more heavily upon the corporation. In short, no logical reasoning dictates that the corporate form of business should be singled out as a base for special taxes.

The type of rate structure now used in corporate taxation is likewise to be condemned. The *rate* of return upon the investment is of much greater significance than the total net return, yet it is the latter which is made the base for taxation. Two corporations might each have a net income of $50,000; in the one case this might represent a return of 20 percent on the investment while in the other but 2 percent. If for no other reason, mere fair play would dictate that these incomes should not bear the same tax, yet in the present method used by the federal and state governments, in the general tax upon corporations, no account is taken of the rate of return.

Injustice is further intensified by the widespread use of progressive rates upon the income of corporations. In the use of such rates it appears revenue is the sole consideration; the fact that a corporation is an artificial person, in which the effects of the tax are felt by the natural individuals who make up the ownership of the corporation, seems to have been overlooked. Assume corporation "A," for example, to have a net income of $100,000, while corporation "B" has a net income of $25,000. With progressive rates the tax upon "A" will be more than four times that upon "B." No consideration has been given to the burden of this tax upon the natural individuals who comprise the corporation. Upon investigation, we discover that corporation "A" has one hundred shareholders of equal amount, while corporation "B" has but five shareholders. In "A," then, each owner has a claim to but $1,000 of the income, while in "B" each has a claim to $5,000. The progressive rate upon the corporate income results in a regressive rate upon the individual. Usually some one can be found to defend almost any proposal in taxation, yet few would volunteer to defend successfully the levy of a smaller rate upon incomes in the higher brackets than upon those in the lower ones.

It is evident that, in the development of our tax system, we have traveled certain roads which are leading us astray from the goal of harmonious relations. Either we must retrace a part of the journey already traveled or we must find some crossroad to return us to the main highway from which we have strayed. Perhaps a crossroad exists, much easier to travel than at first imagined, which will return us to a defensible plan for the taxation of business and individual incomes as the foundation upon which a harmonious system can be built.

Henry George contended that land values are socially created, and in this he was correct. He apparently failed to realize, however, that society has played a considerable part in the creation of all values; that Robinson Crusoe would have met with little success as a merchant, a lawyer, a dentist, or a teacher. His proposal to take the entire amount of one type of socially created value, rent, in taxes, while other types of similarly created values would remain untouched, can hardly be defended. The fact is that society has been a partner in the success of every business enterprise, whether it be conducted by an individual, partnership, or corporation. As a partner in every business enterprise, there is no reason why society should not receive some contribution from business for the part it has contributed to the success of business. Such an exaction in the form of a tax would be but a *quid pro quo* payment upon business; any taxes levied upon persons, *as such*, must be levied in addition to this.

In the administration of such a payment by business enterprises to society, we are at once

confronted by two problems: (1) the calculation of the share which society can rightfully claim and (2) the method by which it can best be secured. Henry George considered rent an economic surplus since it was socially created and since individuals would be just as active in production even if rent were taken by the government. The same concept can be generalized and be made to apply to all production. Just as George designated rent as an economic surplus, likewise the concept of a general economic surplus can be visualized as that part of the social income which goes to any factor of production yet which is not essential to retain or to stimulate desirable economic activity. This might be designated as the economic limit of taxation as distinct from the legal or political limit. As long as the limit is respected, the patrimony of the state will not be impaired, for production will continue in a healthful state capable of meeting the demands of government. This surplus obviously would not always represent the same proportion of the social income. In a society which has reached a high state of development in its transportation and other productive facilities, it would be much higher than in one which still has such development before it. While this surplus represents the economic limit of taxation, it should be emphasized that the actual amount of taxation should be no more than to produce revenue sufficient to finance the justifiable activities of government under efficient and economical administration.

To segregate the exact part of the return to business for which society has been responsible is of course impossible. Likewise it is just as impossible accurately to measure the contribution of individuals through their mental and physical exertions and savings. The administration of such a tax would necessitate arbitrary decisions; for that matter, however, in the administration of what tax are arbitrary decisions absent? Since arbitrary decisions are unavoidable, it may be proper to propose that the primary credit for the success of a business enterprise should go to the individuals connected with it, while the contribution made by society should be given secondary consideration. In accordance with this proposal we should allow a payment to compensate for individual initiative before any return is made to society. Whether an individual be engaged in the activities of a private enterprise or profession, in a partnership or corporation, his services could be considered worth $\$x,000$, which amount would be subject to no *business* tax. As the receipts increase above this, however, an increasingly important contribution should be assigned to society, and payment therefore be exacted at progressive rates as the surplus became larger.

The same logic applies to the net income of business units which has not been assigned to the individual recipients. Those who have risked their savings in business enterprises are more entitled to a return on them than is society. In the administration of the tax, then, a return of x percent should be allowed upon the investment before any recompense is made to society. When more than this percentage is earned, however, society should be accorded a share and should exact payment at progressive rates as the percentage of earnings increases.

Such a method for the taxation of business has a more logical basis for defense than any plan which we have yet used. It is, moreover, administratively feasible, elastic in meeting changing revenue needs, and would take from individuals only a part of that surplus which they had perhaps hoped to receive but had scarcely anticipated. It should be noted, too, that this suggestion in no way disturbs the possibility of taxing the income of individuals, *as such*. Undoubtedly the individual, as distinct from any type of business with which he may be associated in any capacity, should be taxed upon his personal income from services, from investments, or from whatever source derived. The principle to govern in the taxation of personal incomes, however, will be the ability of the individual to pay, and exemptions and rates should be determined on this basis.

The acceptance of the above suggestions as the foundation for our tax structure would offer one way for the elimination of growing friction between state and federal officials. All taxes, of course, must be met from the same over-all social income, and state officials in recent years have viewed with alarm the encroachment of the federal government upon remunerative tax bases. It may not be far afield to suggest that remunerations made to society should go to the federal government, while taxes paid by individuals should remain in the states and localities. The larger social structure is likely to be the contributor to the economic surplus of business, while the personal interests of the individual are likely to conform more nearly to his state and local political units. All levies upon business

enterprises, then, could be made by the federal government, while taxes upon individual incomes, as such, could be left to the states. Each would then have an important separate, elastic, and defensible base for taxation.

Our remaining inquiry concerns the impact of the above plan, defensible from the fiscal standpoint, upon our social and economic structure. It is beyond our province to determine whether any social and economic changes shall be deliberately planned or, if such be planned, whether taxation should be the instrument employed to effect the contemplated changes. Our only contention is that from the fiscal standpoint, if we take economic surplus as an important source of revenue, less friction with the normal course of the economy will be generated than in the use of many other possible bases for taxation. If, in addition, the decision be made to use taxation as a means to effect certain social or economic changes, the use of the economic surplus as a base upon which to make a tax levy is admirably suited to this purpose.

Without arguing the merits involved, the fact is that, in recent years, much emphasis has been placed upon some form of "share the wealth" program, and especially to prevent large inheritances and salaries. If such an end be socially and economically desirable, a more defensible policy would be to "lock the stable before the horse is stolen" than to lock it afterward. It would be more defensible, from the standpoint of the impact upon the economic structure, to prevent the accumulation of large fortunes than to break them up once they have been accumulated. The use of the suggested tax upon business and personal incomes could accomplish this to any predetermined degree.

As another illustration of the possible consequences of this type of taxation, note that it might act as somewhat of a governor in directing our economic system towards business stabilization. We are not here concerned about the causes of depressions, but the essence of a depression is a condition in which productive capacity is far in excess of the capacity to purchase the goods at a price which will meet their cost of production. Stabilization would result through keeping the capacity to produce and the capacity to purchase in balance, and it is towards such an accomplishment that a tax upon the economic surplus could be of some service. For purposes of illustration, assume that a government has been taking one-fourth of the social income in taxes, all from the small-income group, which, if left with them, would have been spent for goods. Another one-fourth of the income is a surplus available for the expansion of a productive capacity beyond that which is needed. Under this arrangement we curtail the power to purchase goods at the same time we encourage expansion of productive capacity, both of which encourage instability. If, however, we should reverse the process and do not take in taxes that which would otherwise have been spent for goods, but take instead that which would have been used for overexpansion of production, we have kept the power to purchase goods and the capacity to produce them more nearly in line; in other words, the system of taxation is directed towards the accomplishment of business stability.

Individuals, both as such and as operators of business enterprises, have looked upon taxes much as they have looked upon the death of a dear friend. They are something inevitable, yet to be deplored and to be avoided as long as possible. This attitude has continued through the centuries, engendered and magnified by the corruption of the tax system to the attainment of private ends, to emphasis upon receipt of revenue at the expense of justice, to the flagrant laxity in the economy and efficiency in the administration of public funds, and to the meager knowledge of the possible meritorious services to be received from governmental activity. Once the members of the social group become convinced that public expenditures can be justified, a receptive attitude has been established within it for the adoption of a defensible system of taxation.

Before much progress can be made towards a harmonious program, a considerable amount of pruning must be applied to what has already developed in the form of constitutional provisions, legislation, and court decisions. The only way to deal with a cancerous growth is to remove it, and we can never hope to secure a healthful and harmoniously functioning tax structure if we retain such a disrupting element, for example, as the present method of taxing business and personal incomes. Those who pay the taxes must be convinced that the method of levy is fair; once this is accomplished, however, a salutary result may be expected in the problem of tax evasion. The great majority of the American people believe in "fair play"; once they are convinced of the fairness of a tax, evasion will not be such a serious administrative problem.

Analysis of the nature of the economic surplus

indicates that it meets admirably the specifications of the type of tax base for which we are in search. To take a part of the surplus in excess of a satisfactory remuneration for personal effort and saving is the nearest approach to the millennium of burdenless taxation. Here is a tax base, moreover, which not only squares with the tenets of justice, but one which can be made to produce large amounts of revenue and which is administratively feasible and politically expedient. Since there is no possible escape from the social and economic consequences of any tax, the use of the economic surplus as an important base for taxation would result in a minimum of friction in the economy. When to these characteristics is added the adaptability of such a tax to effect predetermined social and economic changes, the case for its extended use becomes still stronger. In short, from the standpoints of revenue, justice, and the impact upon the social and economic structure, greater emphasis upon the taxation of the economic surplus is definitely indicated. In fact, posing as a fiscal physician, I make such a prescription in order more nearly to harmonize the social and fiscal aims of taxation.

THE FEDERAL, STATE, AND LOCAL TAX STRUCTURE AFTER THE WAR

MABEL NEWCOMER

Vassar College

(Read February 19, 1944, in the Symposium on Taxation and the Social Structure)

THE interrelations of federal, state, and local tax systems after the war will be concerned with more problems than can be touched upon in this short paper. But the questions that are being asked most frequently today are: Should—and will—the federal government relinquish to the states after the war some of those taxes that it has recently developed for its own use? Which taxes are suitable only for the central government, and which can be successfully administered by the states? And finally, can federal and state governments share the same tax sources without undue hardship to the taxpayer, or co-operate in other ways to insure adequate revenues for the states?

These questions cover the fundamental issues. The local tax structure cannot be disregarded, but it has not been greatly affected by the war, whereas the expansion of federal taxes offers a new threat to the post-war state tax systems. Consequently, federal-state relations are apt to require greater adjustments.

Forecasts are usually, and almost necessarily, based on past trends. Consequently, I shall trace very briefly the development of federal and state tax systems in recent years. This does not require any review of our early history. Only in this century has there been any serious overlapping of federal and state tax systems, in spite of the fact that both governments have always had wide tax powers. Until recently the federal government, on the one hand, limited itself, except in times of unusual stress, to customs duties and a few excises—mostly on liquor and tobacco. The states, on the other hand, had no power to levy customs duties, and they used excises with restraint. More than half of state tax revenues came from the general property tax in 1902. Most of the remainder was derived from special taxes on corporations and other business enterprises, from inheritance taxes, and from liquor licenses. State excises on either liquor or tobacco were unknown.

This century has shown a rapid expansion of tax levies by both the federal and the state governments—the inevitable result of expanding government services; and with this development overlapping taxes have multiplied. By the end of the First World War federal and state levies were in force for corporation and personal income taxes, death taxes, and stock-transfer taxes. By the beginning of the Second World War they had been extended to tobacco excises, motor-fuel taxes, gift taxes, alcoholic-beverage taxes, and payroll taxes. Since the beginning of this war the motor-vehicle license has been added to the list of overlapping taxes. Except for the gift and stock-transfer taxes, all these are levied in two-thirds or more of the states. Three-fourths of state revenues, and an even larger proportion of federal revenues, come from these overlapping sources. The only important independent state taxes remaining are the general sales tax, special corporation taxes, and the general property tax. One of these—the general sales tax—is continually threatened with federal competition. And one—the general property tax—has been largely ceded by the states to local authorities.

Taxpayers and state officials have viewed this growing overlapping with alarm. In fact, the first resolution passed at the first conference of the National Tax Association was a demand that the inheritance tax be reserved for the states. At this time (1907) there were no overlapping taxes, but President Roosevelt had proposed a federal inheritance tax. At succeeding conferences federal taxation of corporate and personal incomes was protested, and objection to federal death taxes was voiced repeatedly. Recognition that the federal government, with its wider jurisdiction, is apt to be more effective than the states in administering taxes has led to state, rather than federal, fear of competition, and consequent protest.

There can be no satisfactory test of the "rights" of different jurisdictions to specific revenue sources, except for legal rights. However, priority in the field, effectiveness in administration, and need are all assumed to give a taxing authority special claims. If the test of priority

is applied, each jurisdiction has encroached on the other's preserve. Death taxes, stock-transfer taxes, motor-vehicle licenses, and motor-fuel taxes were well established state sources before the federal government levied competing taxes. Liquor and tobacco excises, on the contrary, had long been accepted federal sources before the states adopted such levies, and gift and payroll taxes were initiated in the 1930's by the federal government. Both jurisdictions had used corporation and personal income taxes during the nineteenth century, but federal levies were emergency measures, and state levies were few and ineffective. Modern state income taxes are usually considered to have been initiated by Wisconsin, in 1911—two years before the federal tax was imposed, but after the constitutional amendment providing for it had been launched. The federal government was two years ahead of Wisconsin in taxing corporate income. It seems to be almost a draw! Actually, of course, if we were to return to a system of separation of tax sources, neither the federal nor the state governments would accept priority of levy as the test of a proper division of revenues. Need and ability to administer the taxes would be more acceptable tests.

This growing competition for tax revenues is not peculiar to the United States. Other federal states have had a similar tax history. In Canada, for instance, Dominion tax revenues did not overlap those of the provinces and local governments in any important degree until the First World War. The exigencies of war budgets led to the introduction by the Dominion of overlapping corporation taxes and personal income taxes at that time. Sales taxes and stock-transfer taxes (both introduced first by the Dominion) were added to the list of overlapping taxes between wars; and new Dominion levies on inheritances and gasoline, duplicating earlier provincial levies, have come with this war. This leaves the provinces with no independent source of tax revenues. Worse than this, the Dominion has forced the provinces to relinquish all corporation and personal income taxes in return for a fixed pension determined by former yields. Presumably, this is only for the duration of the war.

The Australian federation has followed the same road. Virtual separation of Commonwealth and state tax revenues prevailed until 1910, when a Commonwealth levy on land came into competition with long-established state taxes. This was followed, in 1914 and 1915, by Commonwealth income and inheritance taxes that duplicated state taxes. With the Second World War, the Australian Commonwealth took over income taxes for the duration, forcing the states to withdraw from this field. Here, as in Canada, compensation has taken the form of fixed grants determined by former yields. This is a stage that we have not yet reached in the United States, but it has no doubt given some state officials sleepless nights.

The similarity of these developments is not, of course, accidental. Federal states regularly provide for free trade within their borders. A national economy is the inevitable result. This, in turn, leads to nation-wide controls, and national activities expand. Former local functions are increasingly of national concern. The condition of city streets is still a local affair, but through highways are of national importance. With the increasing mobility of our working population, education, social insurance, and welfare activities are a matter of national concern. Even the police function has growing national significance.

These expanding federal services require expanding national revenues. Moreover, national administration of most taxes is more effective than that of the smaller jurisdictions; for much of the tax base is created by economic activities that cross state borders. Any situs for purposes of local taxation must be arbitrarily assigned. This does not mean that we shall—or should—work for a highly centralized government in the post-war years; only that as long as the centralizing trend continues in industry some increase in centralization of government is inevitable.

State and local governments will still have an important sphere of activity and they must have corresponding financial resources. It has sometimes been suggested that taxes should be assigned to the jurisdiction best able to administer them, regardless of revenue needs, and that the resulting yields should be redistributed, after collection, to the smaller jurisdictions (which would, presumably, be the ones to suffer deficits as a result of such procedure). Professor Seligman suggested such a solution more than thirty years ago, for personal income, corporation, and inheritance taxes, although at the time the federal government was levying only the corporation tax;[1] and similar programs for the replacement of

[1] Edwin R. A. Seligman, *The Income Tax* (New York, 1911): 655–656.

overlapping taxes with a single federally administered tax, to be shared with the states, have been frequently proposed in recent years.

The advantages of such shared taxes are clear. A single, federal tax would be more effectively and economically administered than overlapping taxes, and would bear equitably on all taxpayers —a matter of particular importance when the taxpayers are competing business enterprises. The disadvantages are the difficulty of establishing a basis of distribution that will measure financial needs of the beneficiaries, and the further difficulty of achieving economical and responsible expenditure when the spending authorities are not directly responsible to the taxpayers for these funds. Neither of these difficulties is serious enough to preclude the use of occasional shared taxes. It is only when such funds become a large part of total revenues that these objections to them are valid.

Other solutions are the use of (1) the crediting device now in effect for estate and payroll taxes, (2) grants in aid (which have grown rapidly in recent years), (3) unification in the hands of a single authority of the administration of federal and state taxes on the same base, and (4) division of tax bases between the two authorities. All these may well be employed in some measure in the post-war period, with advantages to state and federal governments on the one hand and to taxpayers on the other. Those who would advocate a single device to solve the problems of multiple taxation are oversimplifying a very complex problem.

It is impossible to set up a model system of federal, state, and local taxation without more knowledge of the probable division of government functions and costs among the different levels of government. Moreover, conditions change so rapidly that it is inevitable that goals will change frequently, and before the earlier goals have been attained. It is possible, however, to indicate some of the trends, and some of the alternatives that are at least worthy of consideration.

Beginning with local governments, the one dependable source of revenue, because it is the least mobile, is real estate. Just because it cannot escape there has been a tendency to overburden this tax base in past years. Excessive burdens have arisen, however, more from unequal assessments, evasion, and exemptions, than from the total amounts collected. The cure lies first in the complete abandonment of such taxes for state purposes; second, in better assessments, perhaps following Wisconsin's lead and extending state controls in this field; and third, in decreased exemptions. This demands an adequate system of payments in lieu of taxes for growing federal and state real estate holdings. It demands, also, the retention of low-cost housing projects in the tax base. Exemption from taxes has offered one of the more obvious ways of keeping rents low, but if low-cost housing is to be developed after the war on the scale that now seems probable, we shall have to choose between taxation of such projects and the abandonment of any important sphere of independent local government. It is quite possible that a single housing project will increase the value of surrounding real estate enough to compensate for exemption. It is not possible that exemption of half of the real estate will double the value of the other half of the real estate, particularly if the taxes on this half are doubled at the same time.

One variation of the real estate tax that merits serious consideration is the adoption of rental taxes. If part of the local taxes is levied on capital values and part on rentals, the total burden on real estate will not be reduced, but the distribution will be different. Because of the tendency for taxes to stick, in some measure, where they are originally levied, such a distribution of taxes would presumably increase the share ultimately borne by occupiers as distinguished from owners. Also, such a distribution of taxes would place a relatively larger share of the total on occupied property and less on unoccupied property.

Large cities can occasionally supplement such levies with other taxes, as New York and Philadelphia are doing. Sales and income taxes are not feasible for the smaller jurisdictions, however, and they only multiply problems of double taxation. Local governments may expect the continuation of grants in aid, and they will doubtless be relieved entirely of some of their more costly functions, such as unemployment relief, old-age support, and other welfare functions. I hold no brief for rapid transfer of functions to the larger jurisdiction, but when it is necessary to choose between state administration and state-supported local administration, I believe that the advantages will usually lie with the former. Local self-government has no virtues unless it is responsible local self-government.

Federal and state tax systems are necessarily more complex than local. Both federal and state

governments are too well entrenched in the field of income taxation to suggest that either should withdraw. Much could be done, however, to avoid inequitable state taxation for those corporations engaged in interstate business, if the states would agree on a uniform system of apportionment of income. This is not simple, of course. Even when states go so far as to accept an apparently uniform formula for allocation of income, they may differ in their criteria for determining the situs of sales. There is no defensible economic basis for such apportionment. Income derived from interstate transactions does not have a local origin. Nevertheless, an arbitrary but uniform allocation is better than the existing chaos.

As between federal and state jurisdictions, duplicate administration can be avoided for both corporation and personal income taxes, even with duplicate taxation. If such procedure is to contribute to simplification and economy, it requires agreement between federal and state governments concerning the definition of income. It does not demand uniformity of rates and personal exemptions. Such procedure could be adopted by agreement between the federal government and individual states without requiring its acceptance by every state. It has been used successfully in both Australia and Canada. If we are to move farther than this on the road toward uniformity, the states must either abandon all claim to income tax revenues or obtain their share from redistribution of federal tax revenues. The latter alternative is, in my judgment, undesirable because the sums involved would probably be large enough to invite irresponsible spending. This was the experience of Germany in the 1920's when the federal government undertook to share personal and corporation income tax collections. Moreover, the German states lost, in the end, because of frequent reductions in their share. This is not an inevitable consequence of such a system, and the fact that our states are more powerful than the German states might well prevent reductions; but it seems probable that such a system would lead to constant bickering over the proper distribution of shares, and that present frictions would only be accentuated.

Death taxes have been a source of particular difficulty in intergovernmental tax problems. Overlapping state taxes have been so serious that extensive reciprocal legislation has been called forth to remedy the situation. Reciprocal legislation has prevented some double taxation, but until there is a uniform definition of domicile the problem of multiple taxation remains, as demonstrated by the Dorrance case. To judge from the literature on the subject, the states have opposed federal invasion of this field of taxation more vigorously than any other tax field. Actually, however, there are few taxes less suited to state administration. In addition to the problem of multiple taxation, competitive bidding for wealthy residents through tax exemption threatened to destroy this source of revenue entirely before the federal credit was established. Federal intervention, far from impairing an important state tax source, has preserved it for those states that choose to employ it. And quite aside from legal tangles, the irregularity of yield of this tax makes it more suitable for the larger than the smaller jurisdiction. The tax on the Dorrance estate increased New Jersey's tax revenues from all sources by one-sixth in 1936. A Du Pont estate yielded Florida ten times the revenues from estate taxes in 1939 that had been obtained from this source in 1938. These windfalls are not, of course, unwelcome. The Du Pont tax saved Florida from a prospective deficit. But these are not the dependable and budgetable revenues that are usually assumed to be appropriate for state use.

For the future it seems desirable to continue the federal credit. The federal tax needs thoroughgoing revision, with simplification and coordination of estate and gift taxes and the provisions for state credits. In the process the crediting provisions might be made more generous. It is particularly desirable that they be extended to the smaller estates, since the yield from these is more stable in amount than from the larger ones, and they are perhaps more local in their nature than the larger fortunes.

In the field of excise taxation, the federal government can probably afford to withdraw from the motor-fuel-tax field after the war, leaving this lucrative source to the states as long as they are primarily responsible for highway building and maintenance. The motor-vehicle-use tax, also, should be abandoned by the federal government. Tobacco taxes, on the other hand, are too firmly established as federal sources to be yielded to the states. If double taxation is to be avoided in this field, either the states will have to give up this source of income or a shared federal levy will have to be introduced. Since it seems probable that the states will need this income, the latter solution is probably the better one. The

revenues involved should not be enough to result in the serious disadvantages of large redistributions of centrally administered taxes.

Liquor taxes should probably continue as independent and overlapping sources. Not only are both jurisdictions dependent on this source for large revenues, but also the element of control involved makes it desirable that independent state authority should not be impaired. Here, and in other areas where duplicate taxation remains, much can be done to reduce inequities and frictions through federal and state co-operation. This will sometimes require revision of tax laws. Often, however, all that is needed is administrative co-operation. Such co-operation as now exists has proved its worth. It is urgent that it be extended.

Most of the concessions noted have been concessions on the part of the federal government. The states, too, have concessions to make. First and foremost, they should accept taxation of interest on state and local bonds under the federal income tax. This is important, first of all, in the interests of equity. Also, it will add to federal revenues at the same time that it equalizes tax burdens. The losses that state and local governments may suffer from higher interest rates have often been greatly overestimated; and in any event the gains for the country as a whole would far offset any individual losses.

If the tax systems resulting from the changes outlined produce inadequate revenues for state and local governments, some extension of grants in aid in the fields of welfare and education may well be resorted to in order to balance budgets.

Many points have been left untouched in this discussion. I believe, however, that these are the more important and more hopeful fields for exploration. Certainly they afford ample scope for the energies of even the most ambitious reformers.

POST-WAR SOCIAL ORIENTATION OF STATE TAX SYSTEMS

WILLIAM J. SHULTZ

College of the City of New York

(Read February 19, 1944, in the Symposium on Taxation and the Social Structure)

STATE governments in 1943 collected $3,900,000,000 in taxes exclusive of payroll taxes, $5,100,000,000 inclusive of payroll taxes.

At the present time these figures may appear relatively small beside the $40,000,000,000 to $50,000,000,000 that the federal government will be collecting under the new tax law. Even if we concede a substantial increase in state taxes after the war—state tax collections have tripled in the past ten years, and the war interrupted a continued upward trend—these increased state taxes will still be a minor fraction of the combined federal-state-local tax budget, for federal taxes will probably never again fall under $20,000,000,000 a year and local taxes will run to some $4,500,000,000 a year.

But the $4,000,000,000 to $5,000,000,000 of the annual state tax bill cannot be dismissed as a trifle. The taxpayers who pay this sum are all too well aware of it as a reduction of their incomes or an increase of their costs. The choice of the tax bases and rates that enter into the state tax systems exercises perceptible influences on our economic structure, and affects significantly the social distribution of our general tax burden.

We have figures for state collections from the various classes of taxes levied. We know that gasoline and motor-vehicle taxes, which were yielding nearly $1,500,000,000 a year by 1942,[1] are the most important class of state tax. Close behind them is the payroll tax, which yielded $1,200,000,000 in 1943. As of 1943 corporation and personal income taxes produced nearly $700,000,000, general sales taxes over $650,000,000, and liquor and tobacco taxes nearly $500,000,000.

We have up-to-date figures on the collections of each state from these sources.

With such full and precise data available, and with the studies that have been made in recent years on tax incidence, it would seem that we should be able to calculate the social distribution of these state taxes—how they burden families at various income levels, how they rest upon farmers and wage-earners and other social groups. But such a calculation for the country as a whole is impossible because of major differences in tax systems as among individual states. It would be meaningless to calculate, on a national scale, the social distribution of the $300,000,000 of personal income taxes, since this would overlook the fact that the burden of these income taxes is distributed in varying proportion, because of rate schedule differences, among the populations of the 33 states that levy personal income taxes, while no state income tax burden is imposed upon the residents of the 15 other states. The same invalidating distortion would enter any calculation of the national distribution of the burden of state sales taxes, or state corporation taxes, or state excises, or state motor-vehicle and gasoline taxes, or state property taxes, or any class of state tax.

Some attempts have been made to calculate social distribution for the taxes of individual states, but with little success. The difficulty here is that, while we have fairly reliable national data on income distribution, such data on state bases are absolutely lacking for most states. Moreover, the burden of certain types of state taxes may rest ultimately in part on non-residents of the taxing states.

I am unable, therefore, to present you with any inductive generalizations upon the present social distribution of state tax burdens, either for the country as a whole or for individual states. This does not mean that we are completely ignorant on this subject, and that our discussion of the post-war social orientation of state tax systems must be launched from complete ignorance of the present social effects of these systems. We can, from our present knowledge of the economic incidence of the more important state taxes, and upon various assumptions of income distribution that would be generally admitted, reach some deductive conclusions as to the social incidence of these taxes.

[1] In 1943, because of reduced motor traffic, the yield of these taxes fell to $1,000,000,000.

Many fiscal authorities deny that payroll taxes are true taxes, and prefer to consider them compulsory insurance premium payments. But state payroll taxes are paid by one group—employers, while the benefits of unemployment insurance are received by a different group—employees. There is sound ground, therefore, for considering these state payroll taxes as true taxes and for tracing the distribution of their burden.

For a time after their initial levy, employer payroll taxes constitute a producers' cost, partly reducing producers' profits, partly affecting prices. In the long run, and particularly where there are classes of labor not brought under the payroll-tax-financed social security system, and hence giving rise to no employers' payroll tax liability, a substantial part of the burden of the tax is shifted back to employees through lowered pay rates or through retardation in any increase of pay rates.[2]

Such part of the state payroll tax as enters the price structure clearly constitutes a regressive tax burden, as does any tax affecting the general price level. And even the part of the tax that in the long run operates as a pay rate reduction must be deemed regressive since, while this reduction may be crudely proportional to payrolls, it does not affect investment income which is received in increasing proportion by the higher income classes. Thus, if we consider state payroll taxes as true taxes rather than as insurance premiums, we have, at the outset, over one-fifth of the total of state tax receipts with a definite regressive influence.

Motor-vehicle and gasoline taxes paid by passenger-car owners have a mixed social incidence. Few individuals in the lowest income classes own cars. But car ownership is common in quite low income classes, particularly in rural regions. Low-income car owners generally own the cheaper, lighter, less powerful styles of cars, which pay lower motor-vehicle taxes. They probably limit their driving more than their better-to-do neighbors, so that their gasoline tax payments are somewhat lower. But such increase of motor-vehicle and gasoline tax burdens by income classes of car owners is not proportional to the rising income scale. Except for the lowest income classes, that are free of this particular tax burden, motor-vehicle and gasoline taxes paid by passenger-car owners are definitely regressive.

The motor-vehicle and gasoline taxes paid by commercial vehicles are a cost item to the concerns that operate these vehicles, and a fairly universalized cost item. As such, these taxes become one in the nexus of price-determining factors. To the extent of their embodiment in prices, they are markedly regressive in their social incidence.

The general sales taxes now imposed in 23 states are also shifted in major part into prices. Various frictions sometimes operate, particularly for a year or so after the imposition of such a tax or an increase of its rate, that prevent full shifting of the tax in all cases. Whenever a sales tax remains as a burden on a business enterprise, it reduces profit income. But these occasional shifting frictions are not of sufficient importance to qualify seriously the generalization that state sales taxes are taxes on personal expenditure and as such are regressive.

Variations in liquor consumption are so much a matter of individual taste that attribution of any social incidence to liquor taxes is rather meaningless. To the extent that liquor drinkers can be grouped by income classes, however, there can be no question but that liquor consumption does not increase proportionately with increase of income. Therefore such social incidence as may be attributed to state liquor taxes is regressive.

Even more regressive is the social incidence of tobacco taxes. There is little difference in annual expenditures for tobacco products by rich individuals and poor individuals. Hence the amount of tobacco tax paid by any individual is much the same whether he has a large income or a small one.

In a recent publication[3] I estimated that, in the long run, and taking capitalization into account, the burden of state and local property taxes rested 25 percent on the possessors of taxable wealth, 15 percent as an increase of commodity prices, 10 percent as a reduction of business profits, and 50 percent as an increase of housing expenditures. The 65 percent of the burden of property taxes that operates to increase housing expenditures and prices has a regressive social incidence. Probably this more than offsets the progressive incidence of the 35 percent of property tax burden that is borne by

[2] William J. Shultz, "Economic Effects of Payroll Taxes," *Taxes:* p. 5, January, 1944.

[3] William J. Shultz, "Economic Effects of the State and Local Property Taxes," *Taxes:* p. 548, October, 1943.

the possessors of wealth and the recipients of business profits.

Not until we come to income taxes and death taxes, which account for only some 15 percent of state tax receipts, do we have state taxes that are essentially progressive in their social incidence. The state death taxes, most of which have been tailored by now to conform with the federal estate tax credit, rising from 3/4 percent to 16 percent, are generally conceded, whether from the viewpoint of the decedent or from that of the beneficiaries, to impose a progressive burden. Most state personal income taxes also have progressive rate schedules. The progression is mild, however, compared with federal standards; while we do have the case of the North Dakota tax, with a rate rising to 15 percent, the normal maximum for state personal income taxes is 6 or 7 percent.

Corporation income taxes are generally assumed to have a progressive social incidence, on the ground that they are borne by the stockholders as deductions of profit available for dividend distribution. But some state corporation income taxes, because of the capital-stock or employed-capital taxes that are associated with them to provide tax payment "floors," may operate to some extent as cost factors and so affect prices. To whatever extent this occurs, a regressive influence enters to offset the general progressive effect of these taxes.

While we cannot calculate any exact or even approximate schedule of the ratios of state tax burden to income for the various income classes, for the country as a whole or even for individual states, the foregoing analysis indicates that the social distribution of state tax burdens, whatever it may be, is unquestionably regressive. Our present state tax systems burden the poorer classes, in proportion to their incomes, relatively more heavily than they do the richer classes of the country.

The pre-war trend of state tax increase is expected to continue after the war, for a number of reasons.

First, there will be the increase normally associated with growth of population. There is no rigid proportional relationship between population growth and increase of governmental expenditure and hence of taxation. Some governmental functions are performed on a decreasing unit-cost basis in relation to population growth. But others must be performed on an increasing unit-cost basis. And population growth, by increasing density of settlement, aggravates various social stresses which must be eased by the establishment of new governmental functions or the disproportionate expansion of existing ones.

Second, we are in an era in which there is a long-term trend of expansion in the services which government renders to its citizens. This is particularly true of the sphere of state government. State functions have been expanding rapidly during the past two decades. This expansion is expected to continue.

Third and most important, pressure for property tax relief in many states is transferring a sizable portion of the cost of government from local units to the states. Throughout our history, except for the period of the last war and at the present moment, real estate has borne a disproportionate tax burden compared with other economic elements. In many areas local real-estate tax rates are pressing against their economic or constitutional maxima, and no additional revenue can be derived from this source to meet vital local expenditure needs. Equity and fiscal necessity dictate the exploitation of other revenue sources as a partial substitute for property taxation. But administrative considerations preclude the development of any other important source of revenue by local units. Therefore the states have been compelled to come to the financial aid of their local units, by taking over functions heretofore exercised by local units, by state aid, and by sharing various state tax revenues with their local units. After the war there will have to be substantial progress along these lines. For this reason, more than any other, state tax revenues will have to be expanded sharply after the war.

This expansion of state tax revenues will occur during a period when federal taxes are being cut from their present high war levels. Local tax revenues during these years will not increase much, if at all. Thus state taxes, after the war, will occupy a more important position in our over-all federal-state-local tax system than they have in the past. The social incidence of the taxes imposed by the states will exercise a greater influence than heretofore on the general social distribution of American tax burdens.

What effect is this anticipated increase of state taxes after the war likely to have on the social incidence of the state tax systems? To answer

this question, we must examine the various classes of state taxes to see which ones will probably account for the post-war increase of state tax collections.

No changes are to be anticipated in the rates of the state payroll taxes. These belong not to the states' general revenue systems but to their unemployment insurance systems. Moreover, their level is determined by the credit allowed for them in the federal payroll tax. No suggestion has as yet been advanced in any quarter that a state might levy a special payroll tax for general revenue purposes and superimpose it on the unemployment insurance payroll tax.

Some increase in motor-vehicle and gasoline tax revenues is probable merely as a result of the increase of motor traffic that may be expected after the war. Some states that still have substantial amounts of highway construction to accomplish, or which will take over some of the roadway functions now exercised by their local units in order to relieve these local units, will unquestionably finance these costs by increasing motor-vehicle license charge rates, or gasoline tax rates, or both. And there are some states in which the motor-vehicle and gasoline tax revenues, at the present rates of these taxes, fall short of fully covering highway expenditures; such states may seek to ease highway claims on their general funds by increasing the rates of their highway taxes. But by and large it does not seem likely that the motor-vehicle and gasoline taxes will contribute in the post-war period to the states' general fund revenues. These levies, rightly or wrongly, have been earmarked to a specific state function—highway expenditure—from the outset, and they are likely to remain so earmarked.

Turning to general-fund taxes, and starting with those that have a progressive social incidence, we find that there is little possibility of increasing the yield of state death taxes, except by increasing substantially the rates on small distributary shares or estates. Large estates are highly mobile. A difference of a few percent in the rates of state death taxes as applied to large estates will cause a substantial number of rich individuals, located in the higher-taxing state, who have retired from active business and who are free to establish residence anywhere in the country, to domicile themselves in the lower-taxing state. Thereby the higher-taxing state loses not merely the death tax revenue involved, but probably also a substantial amount of personal income or personal property taxes during the remaining lifetime of the estate owner. Owners of small estates, however, for the most part have little choice as to the state in which they establish residence. They must locate where their jobs are. They, moreover, are not likely to give thought to how their residence at the time of death may influence the several hundred dollars' tax on their estates. Thus it will be possible for the states to increase the yield of their death taxes by taxing small estates and small distributary shares more heavily than is done now, but this will have the effect of reducing the progressive character of the burden of these taxes.

It is well within the realm of possibility that some of the 15 states that at present do not tax personal incomes will turn to this revenue source after the war. Such a development would tend to reduce somewhat the general regressive distribution of state tax burdens. But if any of the states now levying personal income taxes should seek to derive more revenue from this source, they would be able to do so only by making their income taxes less progressive than they are at present. Many of the recipients of large investment incomes, like the owners of large estates, have considerable freedom of choice as to the state of residence. Any state that imposed a substantial tax on large incomes would simply lose a considerable number of its biggest taxpayers—a sufficient number so that a net revenue loss rather than a gain might be the result of such action. Hence the yield of the present state personal income taxes could be stepped up only by lowering exemptions and imposing heavier rates on small incomes, which do not have the mobility of large incomes.

Corporation income taxes may make a larger contribution to state revenue if some of the 16 states that do not at present use this tax should turn to it after the war. But the 32 states with corporation income taxes at the present time will not be able to do much rate-raising in this tax field. While taxes are not a major factor in determining industrial location, they do exercise some influence in this matter. The states have become acutely conscious recently of the advantages of holding and attracting industry within their borders. Few of them are likely to prejudice their positions by adding to post-war tax burdens on their business enterprises. Indeed, if the recent action of Pennsylvania in providing for exemption of manufacturing busi-

ness from its corporation tax may be taken as a straw in the wind, the post-war tendency is more likely to be in the direction of reducing corporation income tax impositions rather than increasing them.

This cut-rate interstate competition among the states with respect to the taxes with a progressive social incidence does not extend to the regressive sales taxes and commodity and service excises. There is little mobility to consumer retail purchasing. Ordinary day-to-day buying of commodities and services is sharply localized. Price differentials as between states have little effect upon the channels of such buying. The one exception to this generalization occurs where the suburban area of a metropolitan center extends into neighboring states, and even here the margin of influence is relatively narrow.

We must expect, then, that post-war pressure for expanded state revenue will find its major outlet in an expansion of the field of the regressive sales taxes and commodity and service excises. Probably quite a few additional states will be added to the 23 that now levy general sales taxes, generally on retail sales. There is no bar to substantial increases of the rates of the existing general sales taxes. Cigarette and liquor taxes will probably be exploited even more than at present. A number of states are exploring the possibility of post-war amusement taxes with fairly high rates.

Our conclusion, then, must be that the post-war trend in the social incidence of the state tax systems will be in the direction of greater regressivity rather than more progression. This trend will not be the result of any reactionary social objectives on the part of the state legislators. It will be produced by a factor beyond the legislators' control—the effects of interstate competition upon the various kinds of taxes available for state imposition.

What, if anything, can or should be done to counteract this anticipated post-war intensification of the regressivity of state tax systems?

I do not look upon the present regressivity of the state tax systems, or the prospect of their becoming increasingly regressive, as a matter of regret or concern. State taxes are not an absolute social factor. They must be considered in conjunction with local taxes and with federal taxes. Our concern should be with the social incidence of the American tax system as a whole —with the combined social effect of federal taxes, state taxes, and local taxes.

In the combined federal-state-local post-war tax system, federal taxes will far outweigh state taxes, and state and local taxes combined. The federal tax system, even before the war, had a sharply progressive social incidence. With the higher personal income and corporation tax rates that will be unquestionably retained after the war, its social incidence will be still more progressive—too progressive according to some eminent authorities who have written and spoken upon this issue. We can say fairly that an increased post-war regressivity of state taxes will be more than offset by an increased post-war progressivity of federal taxes. We might argue that greater regressivity of state tax systems after the war would be a healthy counterweight to an excessive and unhealthy progressivity of the federal tax system.

But were the anticipated increased regressivity of post-war state taxation to be accepted as *per se* undesirable, it would be possible to prevent or ameliorate it by reducing the need for additional state tax revenue after the war. One method—a questionable one—of accomplishing this would be for the federal government to take over some of the functions heretofore exercised and financed by the state governments. Another would be for the federal government, by an expansion of federal grants-in-aid, to assume a share, or a larger share, in the financing of various governmental functions which would continue to be exercised by the states. A third would be for the federal government to maintain some of its progressive taxes at higher rates than would otherwise be necessary, and share the revenue from these taxes with the states, thereby reducing their need to expand their own tax systems. And a fourth would be an application of the credit system at present used in the federal estate and payroll taxes to the personal income and corporation income taxes, which would enable the states to exploit these progressive taxes to a greater extent without having to fear cut-rate interstate competition. I am not advancing any of these proposals as recommendations, but am merely indicating that there are countermoves, if they were deemed desirable, to the post-war development of heavier and more regressive state taxation.

A POST-WAR PROGRAM FOR THE TAXATION OF CORPORATIONS AND THEIR STOCKHOLDERS

EUSTACE SELIGMAN

New York, N. Y.

(*Read February 19, 1944, in the Symposium on Taxation and the Social Structure*)

WHILE the war has not yet been won, and while there is a very real danger of a weakening of the war effort through diverting too much of our attention to post-war problems, nevertheless in the domestic as well as the international field it is essential that when the war does end, we have programs ready for adoption.

The program for a post-war policy of taxation of corporations and their stockholders which will be advocated in this paper consists of the following four planks:

1. The abolition of all corporate taxes, with the exception of a very small corporation income tax.
2. The adoption of a new method of taxing gains and losses on sales of common stock, in order to prevent undistributed corporate earnings escaping individual income taxes and to allow the individual stockholder the benefit of the corporate losses.
3. The reduction of the individual income tax rates in the higher brackets.
4. A reduction in individual income tax rates on income from new risky enterprises.

I

The abolition of all corporate taxes with the exception of a very small corporation income tax.

When originally adopted, corporation income taxes were in substance a mechanism for collecting the individual's normal taxes at the source, since dividends received from corporations were exempt from normal taxes in the hands of the stockholders. In 1936 this exemption was eliminated, and as a result there has been brought about what is in reality double taxation of corporate earnings. The effect of this is to completely violate all principles of equality of taxation. If a corporation earns one dollar a share and pays a corporation tax of 40 cents on it, and distributes the remaining 60 cents, an individual subject to a 20 percent tax on the dividend pays a tax of 12 cents. Thus in effect he has paid a 52 percent tax on the dollar earned by the corporation for him. If the same business had been conducted in the form of a partnership, he would have paid only a 20 cents tax on that dollar. Even an individual in the 90 percent bracket, by paying 54 cents on the dividend received, would be in effect paying a total of 94 cents as compared to the 90 cents he would pay as a partner. These figures illustrate that in the case of a corporation distributing all of its earnings in excess of income taxes, the effect is that taxes are both greater than the taxes on other income, and also disproportionately high in the case of the small taxpayer.

Similarly, if the same corporation earns two dollars a share and pays the same dividend of 60 cents, retaining the other 60 cents, a 40 percent tax will in effect be paid on the second dollar by both the small and the large stockholder, which is in the one case greater, and in the other case less, than the individual rates of the two stockholders assumed above.

Not only does inequality thus exist if a corporation has earnings, but a similar inequality exists if a corporation has losses. If a business is conducted in partnership form and has losses, such losses are deductible from the other income of the partners. In the case of a corporation, however, the losses cannot be deducted by the stockholders. This not only violates the principle of equality of taxation, but obviously results in much greater injustice to the large taxpayer than to the small taxpayer. If a corporation loses ten dollars a share, a holder of 100 shares in the 20 percent bracket pays $200 greater taxes than he would if he were conducting the business as a partner, whereas a similar stockholder in the 90 percent bracket pays $900 more.

These wholly unjustified inequalities have been accepted by our country in large part because of the fallacious personification of a corporation. We have come to think of a corporation as a profit-making instrumentality which should be taxed as a separate entity. In fact, however,

the corporation is only a legal form through which individuals invest their money. It is the individuals, in the last analysis, who are the ones who alone bear the burden of taxes.

While admittedly an individual stockholder, in contrast to a partner, has little, if any, voice in the decision as to how much of the income of a given year should be held in the business, and while he likewise is free from personal liability for the corporation's debts, these real distinctions between the two forms of organization nevertheless do not justify the existing great difference in tax treatment of stockholders and partners. At most they would justify a tax at a very low rate of approximately 5 percent for the privilege of doing business in the corporate form.

The reduction of the corporate income tax to this rate would therefore appear to be the first step in a sound tax program to eliminate the existing inequalities in the taxation of corporate income.[1]

The effect of this proposal on Government revenues is, of course, a major factor which must be considered. Upon analysis it does not seem as serious as might first appear. In most cases, if corporate earnings before taxes remain the same, the increased earnings after taxes would be distributed to the stockholders, and the loss of the corporate tax would be offset to a large extent by the additional individual taxes, the exact amount of the offset, of course, depending upon the rates in effect for the individual. There might be a tendency for some corporations to accumulate more earnings if the corporate tax were reduced, but even in those cases whatever loss of Government revenue would occur would be largely offset through the adoption of the second proposal discussed below.

The preceding deals only with the corporation income tax and does not deal with the corporation excess profits tax. Obviously, where such a tax exists, the inequalities above shown are increased even further. Furthermore, the present tax is intended to reach war profits only.[2] The excess profits tax on corporations should therefore likewise be abolished when the war is over.

Thereafter an attempt should be made to develop a new type of excess profits tax as an alternative to applying government price-fixing, which now exists in the field of public utilities, to other monopolistic situations such as for example arise from the ownership of natural resources.

The outlines of such a tax were suggested in *The Twentieth Century Fund Survey of Taxation*, prepared under the supervision of a committee of which the writer was a member:

> We do recommend that a strenuous effort be made to apply the tax to unincorporated firms. . . . So far as possible, each business should be taxed separately. Indeed, an effort might be made to isolate and tax the monopoly productions and transactions within a single business, though this may of course prove to be impossible. . . . The rate of return that is called "fair" should vary with the degree of risk. . . .
>
> We have no idea how much revenue can be expected from the kind of tax that we propose. We are not drawn to it primarily as a means of capturing high profits from risky enterprises, or windfalls, or inflation profits. If it were not for the monopoly problem, we should not recommend the tax at all, for we are impressed by its theoretical and technical difficulties. (410)

The current controversy over the so-called renegotiation of war contracts law—which is a misnomer for a 100 percent tax on "excessive profits" from such contracts—has brought out how great these difficulties in fact are. However, if an excess profits tax of this nature is not adopted, the extension of price-fixing is inevitable, and certainly our American experience with rate regulation gives us little confidence in this alternative. Yet we cannot hope to maintain capitalism without finding some solution to this problem.

One other tax on corporations which now exists, namely the capital stock tax and its implement, the declared-value excess profits tax, should unquestionably be eliminated.

II

The adoption of a new method of taxing gains and losses on sales of common stock in order to prevent undistributed corporate earnings escaping individual income taxes and to allow the individual stockholder the benefit of the corporate losses.

If corporation income taxes, other than a small privilege tax, are to be abolished, equality of taxation will be attained in the case of corporate earnings distributed as dividends. However, a fundamental problem still remains untouched, namely, how to effect such equality with respect to undistributed earnings and losses. If this is not solved, undistributed earnings would escape individual income taxes.

---

[1] This was proposed in *The Twentieth Century Fund Survey of Taxation:* 406, 430, 1937.

[2] Carl Shoup, "The Concept of Excess Profits under the Revenue Acts of 1940–42," *Law and Contemporary Problems* **10:** 28, 1943.

This problem has been as baffling to economists as the problem of squaring the circle has been to mathematicians. Various solutions have been suggested.

A. The first is that corporations should be taxed like partnerships. Under this method the stockholder would treat as his personal income his proportionate share of the profits or losses of the corporation.

This proposal is unsatisfactory for three reasons.

Firstly, under the last decision of the Supreme Court dealing with the subject, this proposal would probably be unconstitutional. However, the Court as at present constituted might reverse the earlier decision. If not, a constitutional amendment would be necessary.

Secondly, in the case of a corporation which determined to retain a large part of its profits, this proposal would result in injustice to stockholders who did not have sufficient money to pay the tax on their share of the undistributed profits. If the stock of the corporation had a ready market, the necessary cash could be realized by a sale of part of the stock. However, in the case of corporations with a very narrow market or without any market, such forced sale, if it could be effected, would probably be at a price far below true value, with consequent great hardship to the stockholder.

Thirdly, this proposal would materially complicate the preparation of the stockholder's income tax return and the administration of the tax.

Stockholders would have to be advised by the corporation of their share of accumulated earnings or losses for each year, and would have to include it in their return. In addition, when a stockholder sold his stock, the value of the stock sold would include the accumulated profits upon which he had already been taxed. Accordingly this amount would have to be computed and deducted from the sales price before computing his capital gain or loss. Likewise, if the corporation had sustained losses which the stockholder had already taken advantage of, these would have to be ascertained and added to the sales price. A sale at a profit might thus be turned into a sale at a loss, and vice versa. The resulting complications involved in the preparation of individual returns would be well-nigh insuperable.

In the light of these three objections, this proposal would seem impracticable.

B. Another method of solving the problem which has been widely supported is that of an undistributed profits tax. Such a tax was in effect for 1936 and 1937, in addition to the corporate income tax. In England it is the sole method of corporate income taxation, although it is not usually called an undistributed profits tax.[3] English corporations are subject to an income tax on all profits at the same rate as the individual's normal tax, but when any profits are distributed, the tax paid on such profits is treated as a tax on the stockholder withheld at the source.

This alternative likewise does not provide a satisfactory solution to the problem we are faced with. On the one hand, it would not accomplish complete tax equality between distributed and undistributed profits unless the tax were so high as to force 100 percent distribution of earnings remaining after the 5 percent privilege tax, and on the other hand, even if that were accomplished, the stockholder would still not get the benefit of corporate losses.[4]

Moreover, the suggestion of compelling a 100 percent distribution of corporate earnings is unsound if such distribution must be in the form of cash; the adoption of such a policy, by preventing corporations from retaining any portion of their earnings, would to a very great extent put an end to the possibility of desirable corporate expansion.

Accordingly, various alternative mechanisms have been suggested whereby the distribution of earnings could be made in some other form than cash, but still be taxable to the stockholder and at the same time furnish him with a means of

[3] Shirras and Rostas, *The Burden of British Taxation:* 97 f., 1943.

[4] Kimmel ("Postwar Tax Policy and Business Expansion," *Brookings Inst. Pamphlet* 53: 15, 41, 1943) favors the adoption of an undistributed profits tax similar to that in England with a normal tax rate of 22 percent. This conclusion is justified by the statement: "British experience indicates that a high normal or standard tax collected at the source largely disposes of the undistributed profits tax question" (p. 16, note 22).

The best British opinion, however, is to the contrary: "Successive Finance Acts made special provisions against the practice of using the device of undistributed profits tax on undistributed profits as a means of avoiding the payment of surtax. It is now illegal for companies owned by less than a certain number of persons not to distribute within a reasonable time the income derived by the company to its shareholders. Although these provisions make the intentional avoidance of surtax much more difficult than formerly it would be idle to pretend that the ratio of undistributed profits to dividends, over the general field of companies, would remain the same if these tax differences between the two types of income were eliminated." (Shirras and Rostas, *op. cit.*: 100, note 1.)

paying his tax. One such plan now under consideration is to issue scrip which would be payable in cash in varying amounts to each stockholder depending upon the amount of his income tax; and the balance would be automatically convertible into stock of the corporation. This proposal, while ingenious, is too complicated for possible adoption as well as being open to other objections.

No satisfactory way can be found to avoid the difficulties inherent in an undistributed profits tax, and accordingly it must be rejected.

C. A third solution to the problem which has been advocated is the imposition of a penalty tax upon unreasonable accumulations of earnings. This would be an extension of Section 102 of the Income Tax Law, which imposes a penalty if earnings are withheld for the purpose of avoiding surtaxes on the stockholders. Even in its present form it has proved most unsatisfactory, and it should be restricted to closely held companies rather than extended as suggested.

The reason for this is that it is impossible to devise any precise test of where the line is at which an unreasonable accumulation of profits commences. In effect, therefore, it would mean that some Government agency would have to decide in each particular case to what extent accumulation or distribution of profits was socially desirable, and this would take from private management of business one of its major functions. Furthermore, the administrative problems in creating machinery to provide for prompt decisions by the Government before action must be taken by the directors, rather than decisions by courts some three or four years after the event, as at present, are insoluble. In addition, even if all of these difficulties could be worked out, the major objectives would still not be attained in that the amount of profits reasonably withheld would be free from tax, while corporate losses could not be availed of by the stockholder.

D. Another solution to the problem would be to value all common stocks at the end of each year and to treat the increase or decrease in value during the year as profit or loss to the stockholder.

Wholly apart from the broader question involved in this proposal, namely, the taxation of capital gains, which will be discussed hereafter, this proposal seems unsatisfactory for two reasons.

The one is the impossible administrative task of making such evaluation annually for all corporations. The other is that even if this could be surmounted, the stockholder in a corporation which had no market or only a very narrow market for its stock and who was compelled to sell a part of his stock to pay for his tax on the unrealized gain in a particular year, would in many cases be forced to sell such stock at a sacrifice price far below its true value.[5]

E. We come now to a fifth solution to the problem. Mr. George O. May suggested over twenty years ago that a tax might be imposed on the sale of stock of corporations

> where it might appear that there was a profit which was attributable to the accumulation of undivided profits by the corporation and that the sale was made to avoid the imposition of the tax which would be assessed on such profits if distributed as dividends.[6]

The first half of this suggestion might well be adopted and developed into the following program:

(a) Whenever common stock is sold at a profit, the portion of the profit up to the amount of earnings accumulated by the corporation during the period the stock was held, should be taxed as ordinary income. The balance of the profit, if any, should be taxed as a capital gain.

(b) A corresponding procedure should be adopted in the case of the sale at a loss of stock of a corporation which had accumulated losses.

(c) To enable the stockholder to make the necessary calculations, corporations should be required to inform their stockholders annually of the amount of accumulated profit or loss per share during the year.

(d) In the case of a stockholder buying or selling during a year, the accumulated profits or losses for the year would be presumed to have been incurred proportionately throughout the year, and days within any month could be ignored.

To illustrate this proposal, assume that a stockholder purchased a share of stock in corporation X for $100 on the 1st of April, 1941, and sold the stock on the 1st of July, 1943 for $150. Assume for the sake of simplicity that the corpo-

---

[5] The *Twentieth Century Fund Survey of Taxation*'s conclusion with respect to this solution to the problem of undistributed profits was as follows: "The Committee believes that an unremitting effort should be made to devise some practicable plan by which annual changes in the accrued value of stockholdings can be reflected in the stockholder's income tax returns" (page 479).

The writer is now convinced that this effort can never attain success.

[6] George O. May, *Twenty-Five Years of Accounting Responsibility* 2: 156, 1936.

ration reported to its stockholders that the undistributed earnings per share for each of the years 1941, 1942, and 1943 was $12.

The stockholder in question had a profit of $50 on his share. The amount of accumulated earnings during the period held was $9 for the nine months in 1941, $12 for the full year 1942, and $6 for the six months in 1943, or a total of $27 per share. This amount would accordingly be taxed to him at ordinary rates in 1943, and the balance of his profit of $50, or $23, would be taxed as capital gain.

This proposal would appear to constitute a solution to the problem above presented in that in most cases it would result in undistributed corporate profits being ultimately taxed at ordinary rates to the stockholder, and would correspondingly give the stockholder ultimately the benefit of corporate losses.

What, if any, objections can be advanced to this solution of the problem?

One objection is that since it involves an increase in the tax rate on a part of some capital gains, it is subject to the argument advanced against general taxation of capital gains, namely that it constitutes a deterrent to selling on a rising market and thus tends to accentuate the swings in security prices.

This objection admittedly has some validity, although it has probably been overemphasized. To some degree the desire to avoid the payment of tax is without rational basis, for the tax probably must be paid sooner or later if the profit is to be taken, but under certain circumstances the imposition of a tax on capital gains undoubtedly is a justifiable reason for postponing selling. However, since no tax program can be devised which is perfect, this disadvantage must be weighed against the other advantages of the proposal.

Another argument against taxing capital gains generally as ordinary income is that, since justice would then require capital losses to be deductible from ordinary income, the Government's revenues would fluctuate widely and be increased in boom periods and be decreased in depression periods. Nevertheless within limits this result is not undesirable. Boom periods are the very time when the Government should endeavor to increase the yield from taxes and reduce its debt, while during depression periods it should relieve the taxpayer, thus increasing his purchasing power and thereby lessening the deflationary effects of the depression. Yet if capital losses of all types of assets were allowed as deductions from ordinary income, the yield from the income tax might in periods of depression become wholly inadequate.

However, the suggestion under consideration, of limiting this treatment to that portion of the capital losses of common stocks which represents accumulated losses of corporations, will have a far lesser effect in reducing revenues. Consequently it would appear that this objection to the proposal under consideration is not a very serious one.

A third objection to taxing capital gains generally is that they are affected by a change in the value of the dollar. If an asset cost $100 and is sold for $200, but the value of the dollar has been cut in half, there has been no real gain; the apparent gain is purely illusory. This difficulty might, of course, be overcome by taking into consideration changes in the value of the dollar, although the administrative problems involved in this are obvious.

There is, however, another aspect of this objection. As a result of inflation real loss occurs to holders of fixed obligations. If a bond cost $100 and is sold for $100, and if the value of the dollar is cut in half, the bondholder has really had a loss of $50. Yet it would seem impossible for any tax system to allow the deduction of such losses from ordinary income; the effect of this would probably be, even if similar gains were included, to completely disrupt Government finances during inflation periods. Any system of taxing capital gains generally would require some solution of this problem.

However, a change in the value of the dollar would not create any difficulty under the limited proposal under consideration, which is applicable only to the portion of capital gains and losses on common stocks attributable to accumulated earnings or losses. Therefore it is not an objection to such proposal.

A fourth argument against the general taxation of capital gains is that they do not constitute true income. Whatever validity this argument may have with respect to the taxation of capital gains in general, it likewise would not seem applicable to the limited proposal under consideration, of taxing capital gains to the extent that they represent accumulations of income.

A fifth objection to the proposal is that under the present tax law, by holding the stock until death, the taxpayer would avoid paying the income tax on the undistributed earnings. This

objection can and should be met by an amendment to the law, which perhaps for constitutional reasons might have to be part of the estate tax law, imposing a tax equal to that which the stockholder would have paid under the proposal if he had sold the stock the day before his death at a price equal to its value as determined for estate tax purposes.

Another objection is that of the practical difficulty in the annual furnishing of the necessary data by corporations to their stockholders. This is, however, surmountable. In the case of corporations whose fiscal year is not a calendar year or in those cases where the data would not be ready by March 1, the corporations could be required to make an estimate which would be final as far as the stockholder was concerned, and any error in such estimate would be added as an adjustment to next year's figure. Likewise, in the case of the ordinary corporation, any correction of income made by the Treasury Department upon audit would be added as an adjustment to the next figures reported to stockholders. In this way the stockholder's task would be simplified, although some minor injustices might result.

There is a final objection that the proposal does not produce complete tax equality and permits undistributed earnings in part to escape individual income taxes. This is true. One way in which this might happen would be that stockholders might hold stock appreciating in value for many years, perhaps until their death, and thus, by postponing the payment of the tax, might in effect be borrowing money from the Government without paying interest. It would be obviously impracticable to work out any method of payment of interest, but it should be noted that this advantage would be offset in such a situation by the fact that the stockholder would presumably pay higher rates if he waited to take his profit all in one year. Another situation in which complete tax equality would not be attained would be that of a stockholder who sold his stock for a profit less than the accumulated earnings upon his shares, and thus completely escaped taxation upon a part of such earnings. This could be avoided by taxing the accumulated earnings when the stock was sold, irrespective of the amount or even existence of profit, but such a suggestion is open to other objections, including that of unduly complicating the individual's income tax return.

However, no solution to the problem of undistributed corporate earnings is perfect, and the solution proposed would appear not only preferable to any other suggested alternative, but also far preferable to the present method of corporate taxation which, apart from section 102, makes no attempt ever to tax undistributed earnings as ordinary income to the stockholder.

The first two planks in the suggested program would thus appear together to constitute a method of achieving the maximum tax equality which is practicable between stockholders of corporations and other recipients of income.

III

The reduction of individual income tax rates in the higher brackets.

Income tax rates in the higher brackets have reached the height where they materially reduce the incentive to work. If mankind were actuated entirely by unselfish motives and were not influenced by the rewards for activity, this would not be the case. At the present stage of man's development, even though selfish motives are only a part of the considerations which activate him, as is shown by the example of the dollar-a-year men working for the Government during wartime, nevertheless the inducement of financial reward is a material factor in bringing forth his greatest energies. When income tax rates reach 93 percent at $200,000, the inducement to further activity has been for all practical purposes eliminated.

If the two proposals suggested are adopted, they will result in more income being brought into the higher brackets, which will thus increase the effect of the existing rates in destroying the incentive to further activity. Consequently the necessity of reducing rates in the higher brackets will be increased.

How can one decide what the exact rates should be? When progressive taxation, as contrasted to proportional taxation, was first suggested, it was opposed by most economists. M'Culloch expressed the early nineteenth-century view when he wrote: "When you abandon the plain principle of proportion you are at sea without rudder and compass, and there is no amount of injustice that you may not commit."[7]

The principle of progressive taxation is now universally accepted as justified by the doctrine of ability to pay, but no formula has been found for fixing the rates. Professor Seligman states:

[7] Quoted in Sir Josiah Stamp, *The Fundamental Principles of Taxation:* 40, 1936.

While the usual arguments advanced against progressive taxation are almost entirely destitute of foundation, it is not a simple matter to decide how far or in what manner the principle of progression ought to be actually carried out in practice. Theory itself cannot determine any definite scale of progression whatever.[8]

Since there is no exact method of determining what the progressive rates should be, a decision must therefore be reached by attempting to balance the need of revenues, the principle of ability to pay, and the importance of retaining incentives to activity. Obviously, in balancing these diverse factors, different individuals will reach different conclusions. To the writer peacetime income tax rates graduated progressively up to a maximum of around 50 percent in the top bracket would appear to constitute a reasonable solution. The constitutional amendment now being advocated, limiting such maximum to 25 percent, would appear to be going too far.

IV

A reduction in individual income tax rates on income from new risky enterprises.

Even if income tax rates are reduced as suggested, so as to increase the financial incentive for the individual to work and thereby earn and save, there still remains another problem, namely, that of promoting investment in risky enterprises.

One result of progressive taxation which is not commonly appreciated is that even in the lower brackets it tends to deter risk-taking by in effect changing the gambling odds. Inasmuch as the first two proposals made above would result in subjecting to progressive taxation corporate earnings now not subject thereto, the impact of progressive taxation would thus be increased and this problem rendered more acute.

Suppose, for example, that an investor estimates that with an equal investment he has one-tenth the chance of succeeding in venture A, where success will bring him $50,000 income in one year, as in venture B, where success will bring only $5,000 income in one year. The two ventures thus appear to him equally attractive apart from the income tax. Under the income tax at present rates, however, assuming that his other net income is $10,000, the return from the risky venture A will be cut to $16,680, whereas the return from the comparatively safe venture B will be cut to $2,690. The risky venture, instead of bringing in ten times as much as the other if successful, will then bring in only six times as much.

Obviously this change in the odds will result in the investor preferring the less risky venture. In view of the importance of the development of new kinds of industries, both in furnishing employment and in stimulating the economy as a whole, this result is most undesirable.

Accordingly, it is suggested that a provision be inserted in the tax law, whereby investments in the form of common stock in new risky enterprises should, for a period of time, be granted preferential tax treatment. The working out of the details of such a proposal would involve decisions as to three matters:

First, what is to be the test of a risky enterprise. It would not include all new investment; the construction of a new plant to increase the capacity of a company with an established business, such as a public utility, should not come within the test. The most practical solution would appear to be to define a risky enterprise in general terms and then to provide for an application to the Treasury Department in advance for a ruling in each particular case, which, when made, would determine the matter finally.

The second question is the length of time for which preferential treatment should be granted. A minimum period of five years and a maximum of ten years are suggested as reasonable.

The third question to be decided is the extent to which preferential treatment should be given. It is suggested that one-half of the dividends paid during the preferential period, together with the same percentage of capital gains, be free from income tax. This mechanism would be the simplest from the point of view of the taxpayer in preparing his return.

The degree of preferential treatment above set forth is merely a suggestion, and it might well be considered desirable to grant a greater preference in order to reduce the deterrent effects of progressive tax rates on new risky investments.

Would it be desirable to go further and stimulate not merely risky investments but also all new investments or even possibly all savings? The first could be accomplished by extending the preferential treatment to all new investments, and the second could be accomplished by, in addition, reducing the income tax on that portion of the income which is saved, which is the same

[8] Edwin R. A. Seligman, *The Income Tax:* 33, 1914.

thing as levying a higher tax on that portion of the income which is spent.

Such a program, it will be noted, is the opposite of the provision in our tax law, just repealed, which granted to a very limited extent a preference to so-called earned income, and it is likewise opposed to the views of those who advocated the 1936 undistributed profits tax on the ground that it would tend to decrease corporate savings and put more money into the hands of stockholders for the purchase of consumption goods.

While during wartime conditions such a program would be desirable, for the post-war period it would not seem wise to adopt it, at the present stage of our economic knowledge, for there is too wide a difference of opinion among economists as to whether or not the United States has reached a stage in its economic development where savings are already accumulating too rapidly or, in the alternative, where productive capacity is already being increased too rapidly through new investments. To the writer it appears that, at least as to productive capacity, further increase is desirable and that general overproduction can never exist. However, more study of these difficult questions seems necessary before a decision can be reached as to whether or not taxation should be used to promote or retard savings or investment beyond that involved in the preference to risky investments. The dangers of making the wrong decision because of insufficient knowledge make caution a preferable course.

Even without granting any preferential treatment to savings, the adoption of the third proposal above set forth, namely the reduction of rates in the higher brackets, would of itself increase the incentive to save, and the first proposal of reducing corporate income taxes to 5 percent might tend to increase corporate savings.

CONCLUSION

The four proposals above set forth constitute an integrated program consisting of the substantial abolition of corporate taxes and the changes in our tax laws made necessary thereby. They do not purport to cover the entire field of possible changes in our Federal income tax laws. The difficult question of taxation of capital gains generally has only been discussed incidentally without attempting to formulate a program. Many other important matters have not even been mentioned, such as the exemption from Federal income taxation of state and municipal bonds issued hereafter. Obviously these subjects cannot be dealt with within the limits of this paper.

Furthermore, the above program is, as has been stated at the beginning, proposed for adoption only after the war has been won and when the productive energies of our country will again be turned back to the manufacture of articles of peace instead of munitions of war. When that time comes, it would be desirable that the program should be adopted in its entirety rather than in gradual steps. Whether that will be possible, however, depends upon the conditions which will then exist, such as the extent of our Federal debt, the size of our national income, etc. No one can today prophesy what these will be.

Finally, as already recognized, the suggested program is subject to various objections, some of considerable weight. This is, however, unavoidable in any tax program. As has been so aptly expressed in the much quoted paraphrase of Pope's couplet:

> Whoever hopes a faultless tax to see,
> Hopes what ne'er was, or is, or e'er shall be.[9]

[9] Quoted in Stamp, *op. cit.*: 215.

CATAWBA POTTERY-MAKING, WITH NOTES ON PAMUNKEY POTTERY-MAKING, CHEROKEE POTTERY-MAKING, AND COILING

VLADIMIR J. FEWKES *

University Museum, University of Pennsylvania

(*Read by title November 22, 1941*)

CONTENTS

Introduction 69
Methods and techniques of Catawba pottery-making . 72
Acquisition and treatment of raw material 72
Natural conditions and supplies 72
Clay pits 73
Preparation of paste 73
Process of manufacture 75
General preparations 75
Building technique 77
Modeling 77
Segmental building 78
The uninterrupted process 78
The ring variant 79
The circuit variant 79
The sectional process 81
Molding within a double form 81
Forms 83
Surface finish 86
Decoration 88
Drying process 88
Firing 90
Post-firing treatment 93
Time duration of the potter's tasks 94
Technological miscellanea 96
Cherokee pottery-making 97
Pamunkey pottery-making 103
Some historical retrospects 105
Concluding remarks and comments 107
Addendum: coiling 110
Bibliography 122

INTRODUCTION

THE contemporary descendants of the Catawba Indians, a remnant of the Eastern Siouan-speaking stock, live on a state reservation near Rock Hill, York County, South Carolina. The "nation," as the natives like to refer to the reservation group, preserves several prominent survivals of aboriginal culture. Particularly noteworthy among these is pottery-making. Since 1884, the date of the first written record of the craft by a dependable observer (Palmer's notes, *vide infra*), Catawba women have been repeatedly mentioned as skillful potters. Collections of pottery representing the interim reveal unmistakable uniformity in style and in technical details. Moreover, the modern product has very close similarities, if not full analogies, among archaeological remains found at sites known to have been occupied by the Catawba in early contact times or even before the whites appeared. This circumstance renders the study of the Catawba potter's craft particularly attractive from the culture historical point of view. The usefulness of such an inquiry may be profitably increased by considering the status of pottery-making among the Cherokee of the Smoky Mountains in North Carolina, and among the Pamunkey of Tidewater Virginia. These groups still practice the craft along old traditional lines, and both have been exposed to Catawba influences through intermarriage and intercommunication.

The purpose of this publication is to examine contemporary Catawba pottery-making; to investigate its history by projecting retrospective inquiries as far back as positive evidence permits; and to consider, strictly objectively and within practical limits, pertinent comparative and supplementary data. The notes on coiling became an inevitable addition with which to approach the task of classifying construction in hand-made pottery.

In the United States, the study of either archaeological or ethnological pottery has been coordinated into a rational discipline; currently, several serious and well-qualified investigators are pursuing researches which promise further advancement. The pioneering efforts in the field must be credited to the late Dr. W. H. Holmes, who over 50 years ago began to publish his studies, of which several should ever enjoy the

* Owing to the death of Dr. Fewkes on December 11, 1941, after the completion of this manuscript but before the arrangement of the figures Mrs. Lucile Serrem Paterson, who had worked with Dr. Fewkes on the manuscript shortly before his death, arranged the figures and inserted references to them in the text.

reputation of classics. It is in point to recall that his studies were continental in scope, that they embraced archaeological and ethnological evidence alike, and that they concerned themselves also with lithic, textile, and other domestic industries. Perhaps some of Dr. Holmes's shortcomings may be attributed to the inevitable dangers inherent in such widely spread and diversified work. In any case, he covered several areas in considerable detail, and his contemporaries or immediate successors carried on quite successfully. In the East, several accounts of surviving native pottery were prepared, and some attempt was made to connect certain modern wares with archaeological material. The Mississippi Valley received little attention from the time Holmes published his famous "Aboriginal Pottery" (1903) until Federal Government appropriations of recent years helped to develop extensive exploration programs. In the Southwest, however, favorable circumstances stimulated pottery studies and attracted many specialists. Dr. A. V. Kidder may well be called the Nestor of modern research in Southwestern pottery. His acumen, foresight, and profound scholarly attitude, coupled with rich field experience, rendered him eminently fitted for the mission. This was to prepare the ground for, and to initiate the practice of research in, pottery technology. As a part of Dr. Kidder's program, Dr. C. E. Guthe published his splendid "Pueblo Pottery Making " (1925), which is replete with minute details of manufacturing steps. Then appeared Miss A. O. Shepard's "Technology of Pecos Pottery" (1936), as part of Dr. Kidder's second volume on Pecos pottery. Miss Shepard describes the aims and methods of the research, examines critically the glaring errors committed in subjective interpretation, and presents convincing proof of the practical and indisputable value of technological pottery analysis. Such work is entirely scientific, devoid of personal reckoning, and fortified by standards derived from repeatedly demonstrable principles.

Thus the study of pottery is now facilitated by excellent descriptive and detailed analytical reports based upon a thoroughly objective approach. It is this quality that distinguishes the modern work from so many earlier studies which, however well intentioned, are not always free from subjective treatment.

The value of scrupulously observed objective methods in pottery studies is the more pronounced in view of the technicalities of the subject. Technological work pursues precise determinations by employing various means of investigation proved dependable in other endeavors of cognate aims. The analytical data are compiled not necessarily for classificatory purposes; rather they provide indices of qualitative and quantitative properties of the analyzed specimens. The usefulness of such data, as against the pitfalls of subjective deductions, is quite obvious. There is no excuse for speculation in such matters as nature and sources of clays and temper, chemical and physical composition, or optical properties; and one can also be fully objective in dealing with conquests, dates, diffusional trends, migrations, trade contacts, etc. If the positive evidence at hand is insufficient to yield the necessary data for answering these questions, no attempt is made to substitute opinions. That pottery in itself, no matter how well analyzed, has its own limitations in culture historical investigations, is readily admitted. The signal factor to stress is that studies of pottery in general, excepting aesthetic values, lend themselves to a full application of the principles of scientific procedure; technological analysis,

MAP 1. Approximate locations of Cherokee and Catawba nations. Shaded areas: territories in eighteenth century. Dots: present locations of East Cherokee and Catawba reservations.

as convincingly demonstrated, operates with precise scientific methods. And yet, irrespective of the remarkable refinements reached in technological pursuits, of the splendid body of factual data compiled, and of the standards established, the laboratory work is only one part of the picture. For it is equally interesting and important to know how a given pottery was actually manufactured. In this regard, the scope of technological endeavors may be most profitably enlarged by specific detailed studies of contemporary pottery-making among aboriginal survivals. In the United States a very fruitful field exists for such investigations, with several Indian groups still practicing the craft under traditionally retained principles of aboriginal methods and techniques.

My initial field work in the subject matter here treated dates back to February, 1929, at which time it was my privilege to accompany Professor Frank G. Speck to the Catawba reservation.[1] Favorable circumstances enabled me to observe, even during the first visit, virtually all the different processes and procedures practiced by the potters. Subsequent observations and comparative studies extended the intermittent inquiries into August, 1941, at which time I last visited the Catawba (and also the Cherokee). The sum total of the successive investigations corroborates the initial findings in principle. The most recent field work, naturally, was directed toward a general collation and especially toward an understanding of the contemporary status of Catawba and Cherokee pottery-making.

The population of the Catawba reservation, as estimated by Chief Sam Blue in the late summer of 1941, numbered about 260 souls, representing some 60 families. With the aid of Mr. Irving Brown, 33 families were counted as pottery-makers; that is to say, at least one woman in each was deemed to be fully acquainted with the native processes. The seasonal absence of the younger women from the reservation rendered an accurate check rather difficult if not impossible to obtain, but it seems to be a fair estimate to state that some 70 female members (more or less) of the "nation" know how to make native pottery. In any case, the chief informants from whose work and information the data here presented have been drawn, and the most dexterous performers, were the following (status of August, 1941):

Lula Beck	Rachel Brown	Georgia Harris
Doris Blue	Fannie Canty	Mary Plyler
Louisa Blue	Eliza Gordon	Arzada Sanders
Edith Brown	Sally Gordon	Lillie Saunders (now Bryson)

Despite modern influences, the task of the Catawba potter remains unaffected; the craft is based on the principles of an ancient technique and certainly appears to be neither dormant nor decadent. A distinct consciousness, and indeed pride, are clearly manifested by the contemporary artisans, whose skill and aesthetic sense reflect a deeply rooted control of their endeavors. Local archaeology, although very unsatisfactorily known thus far, supports the view that much of the aboriginal technique is being traditionally retained. The recent archaeological field work of the University of North Carolina, under the direction of Joffre L. Coe, has helped to identify several historic sites. Among them are early eighteenth century locations, long since abandoned, which reliable sources ascribe to the Catawba. The pottery found at such sites includes mottled polished ware which in construction, surfacing, and firing closely resembles the modern Catawba product.

[1] Acknowledgments:

It gives me pleasure hereby to express to Professor Speck my gratitude for his counsel and advice.

The most significant segments of my studies resulting in the present publication became possible through the following generous grants. In 1939 and 1940, the Society of the Sigma Xi, Pennsylvania Chapter, favored me with funds toward specific technical work. A grant from the American Philosophical Society in 1941 enabled me to pursue further field work and to complete the manuscript.

To Mr. Donald Horton, Yale University, Mr. Frederick R. Matson, Jr., University of Michigan, Mrs. Lucile S. Paterson, University Museum, University of Pennsylvania, and Miss Anna O. Shepard, Carnegie Institution, I am greatly indebted for reading either provisional or final drafts of the manuscript and for many valuable suggestions. Mrs. Paterson also helped in proofreading.

During some of my field work, I had the pleasant company of Mr. Joffre L. Coe, University of North Carolina, whose knowledge of local culture history and skill in photography were very helpful.

To the University Museum, University of Pennsylvania, I am particularly obliged for the use of its technological laboratories as well as for certain photographs and drawings. The Peabody Museum, Harvard University, favored me with a number of photographs. Specific acknowledgments are stated in appropriate explanations of the plates. The photographers were Reuben Goldberg, University Museum; Fred P. Orchard, Peabody Museum; Joffre L. Coe, University of North Carolina. The drawings were prepared by Mr. J. S. Benditt of the University Museum.

To my wife, Berenice, I owe profound gratitude for sustained help in the preparation of the manuscript, and for the inspirations derived from her untiring interest to aid me in every possible respect.

Pottery-making among the Catawba is essentially a woman's calling. Occasionally men participate in the digging of clay in natural deposits, a task which often calls for considerable labor, but with the completion of this process their share is usually finished. The various steps of manufacture appear to be a common tribal property on the reservation, and the craft is handed on more or less through family lineage. Sometimes newcomers are instructed by seasoned artisans.[2]

Naturally, individual dexterity is often responsible for certain deviations from the otherwise well-established standards. On the whole, however, the technique is essentially uniform throughout the reservation, and with negligible exceptions the final product differs very little from household to household. Yet, individual potters are able to recognize quite readily the results of their own labor even if their ware be mixed with that of others, although no individual identification marks are used.

The pottery-making season normally lasts from early spring to late fall. Frost is to be avoided, but even in the winter some activity often continues indoors. In the summer, drying is naturally facilitated and, usually, the livelier marketability of the product increases demand. While some seasonal fluctuation in production thus results, the natural limitations imposed by weather conditions are not a very serious drawback to pottery-making throughout the year.

The contemporary Catawba have no tradition regarding the origin of their pottery. No suggestions of a likely explanation are known from their myths and lore, nor are there traces of a record in extant sources which might elucidate this point. The following creation belief, however, is of some interest in this relation, although not exclusively a Catawba occurrence. "In the beginning the creator modeled a man from clay and proceeded to fire his product. The fire was too low and when finished the figure was very light in color. That was the white man. In a second attempt, too much fire and heat scorched the figure—that was the black man. The creator then tried a third time. He gave much care to the fire and arrived at a perfect result—the red man." [3] The absence of a native explanation of the origin allows very little in the way of tangible deductions. Mrs. Sally Gordon, confronted with a request for an elucidation regarding her own conception in the matter, responded somewhat as follows: "Really, I do not know. But I have heard it said that our people knew how to make pots and how to grow corn since a very long time ago." Other people at the reservation declined to venture on any conjecture. The question of the origin of Catawba pottery remains open; indeed, it appears to be a moot one.

My purpose is to present an account of the pottery-making technique observed at the reservation and of the general information furnished by the natives upon specific inquiries.[4] The description follows the procedure employed by the potters, and the various steps are arbitrarily arranged under eight headings:

Acquisition and treatment of raw material
Process of manufacture
Forms
Surface finish
Decoration
Drying process
Firing
Post-firing treatment

METHODS AND TECHNIQUES OF CATAWBA POTTERY-MAKING

ACQUISITION AND TREATMENT OF RAW MATERIAL

NATURAL CONDITIONS AND SUPPLIES

Within the physical environment of the Catawba there are no serious handicaps to finding adequate beds of clay suitable for the needs of

[2] The case of Mrs. Lillie Bryson (formerly Mrs. Saunders) presents an interesting example in point. Lillie, originally a descendant of the Cherokee (born in northern Georgia), married Joseph Saunders on the Catawba reservation. At the time of her arrival she knew little about pottery-making. Yet within a very short time she acquired, largely through the teaching of Mrs. Sally Brown, a thorough knowledge of the local craft. Her products fully conform to Catawba standards. After the death of Joseph Saunders, in 1930, Lillie removed to the Cherokee reservation (Swain County, North Carolina), and married Saunook, an officer of the Cherokee tribe. At the present time she is Mrs. Bryson, and lives at Ela, some 6 miles from the village of Cherokee. She continues to make Catawba-style pottery.

[3] The Catawba do not associate this myth with pottery origins. (I may add that a practically identical version was told me by a Delaware from Oklahoma.)

[4] Catawba linguistic equivalents for the clays, implements, tools, etc., are not included here; the reader may find these in Speck (1934; 47–48 (clay-eating) and 70–72 (pottery-making)) and in Harrington (1908*a*: 402 ff.). (Some of the terminology published by Harrington (1908*a*) requires revision.) For literature on Catawba pottery, see Holmes (1903) and Harrington (1908*a*). Pennypacker (1937*a*) is to be dismissed because of its inadequacies and amateur quality.

the potter. Exposed sides of hills and ravines, hilltops, banks of streams, and occasionally eroded surfaces, are relatively easily accessible. In these, the women potters, usually aided by men, locate the potential sources to be exploited. Under these favorable circumstances a ready supply of clay is always close at hand. Abundant raw material may be obtained within the limits of the reservation or near it.[5] There are no restrictions on the exploitation of the clay beds. Irrespective of land ownership and, apparently, notwithstanding possibilities of trespassing, clay-bearing deposits are free for public use. Some clay beds are situated on land belonging to Negroes and are being exploited by the Catawba without any compensation.

CLAY PITS

Apparently because of the labor involved, the actual digging of clay is done by men, but women, *i. e.*, the potters, attend to the immediate sorting and picking at the source of supply. At times the pits are carefully covered [6] to protect them from rain washes, and often they are concealed and guarded.[7] A given pit need not be exploited for a long period of time.[8] The location of the water table, the thickness of overlying strata, the accumulation of ground or rain water, and general accessibility, are among the decisive factors in this regard. The natives concentrate on a pit, as Harrington points out, "until it becomes troublesome to keep free from water, then abandon it and begin another one nearby." [9] Sometimes the clay beds are situated several feet underground and require the removal of a considerable amount of material to be reached. The digging is usually done with the aid of agricultural implements, although in certain cases it is possible to procure the clay by mere hands. As the bulk is being heaped beside the pit, "impurities" and undesirable particles are removed when they come to view, and the suitable clay is packed for transportation. Sacks,[10] baskets, or boxes are used for this purpose. Two qualitatively different types of clay are utilized: (*a*) the so-called *pan* clay, often also referred to as "blue clay," which is a relatively dry and compact, coarse-textured variety, containing a natural admixture of sand and usually mica; (*b*) the so-called *pipe* clay, fine in texture, somewhat stiff, relatively moist, and wellnigh free of sand, yet often containing minute particles of mica. The two kinds are found in different beds, and in separate deposits, and in the raw state are always stored separately, either "dry" (in sacks, etc.) or wet (in buckets, etc., moistened with water). The pan or blue clay is most abundant in river bottoms and gullies, and appears to be sedimentary. The pipe clay, on the other hand, is most common in elevated locations, and appears to be residual. As shown by laboratory tests, the pan clay has an average linear shrinkage of 4 percent, and the pipe clay has an average shrinkage of 2 percent.

Chemical analyses reveal a high percentage of ferric contents in both the pipe clay and the pan (blue) clay. The actual percentages vary with individual locations. The several samples obtained by me and qualitatively analyzed have shown the following proportions of iron oxide:

Pan clay	Pipe clay
Sample 1: 11.92 percent	Sample 1: 8.83 percent
Sample 2: 10.40 percent	Sample 2: 9.35 percent
Sample 3: 11.67 percent	Sample 3: 10.04 percent

The high percentage of iron helps to explain the red-burning tendency of the two clays.

PREPARATION OF PASTE

The initial treatment in the preparation of the paste consists of pulverizing the clay by vigorous pounding with a cylindrical, double-headed wooden pestle, 0.85 m. to 1 m. long and about 0.08 m. to 0.12 m. in diameter (fig. 1*h*).[11] The raw material is crushed first, and then either sifted through a household sieve, or screened through a piece of window screen. This is done either upon a board foundation [12] or on some coarse textile, and facilitates the removal of objectionable matter which is exposed as the clay is

[5] Harrington (1908*a*: 402) states: "Three mines of pan clay are known on and near the reservation, and five of the pipe clay." This number, however, does not express the potential sources in the locality.

[6] *Cf.* Speck (1934: 71): "When you depart 'from digging clay' put some earth back in the hole to cover it."

[7] Even in 1941, I was asked to remain in the car while a sample of pan clay was being obtained for me by a member of the reservation.

Cf. Harrington (1908*b*: 224) for the significant observation among the Eastern Cherokee: "After Ewi Katalsta had dug her clay from a bed on Soco creek, the exact location of which she did not seem inclined to reveal . . . ."

[8] *Cf.* Speck (1934: 71): "For a long time clay has been dug, now the hole is big . . . ," *i. e.*, ready to be abandoned.

[9] Harrington (1908*a*: 403); for an illustration of a small pit, see his plate XIX*a*.

[10] Harrington (1908*a*: 403); Speck (1934: 70) specifies a bag.

[11] *Cf.* also Harrington (1908*a*: pl. XIX*b*).

[12] Harrington (1908*a*: 403 and pl. XIX*b*).

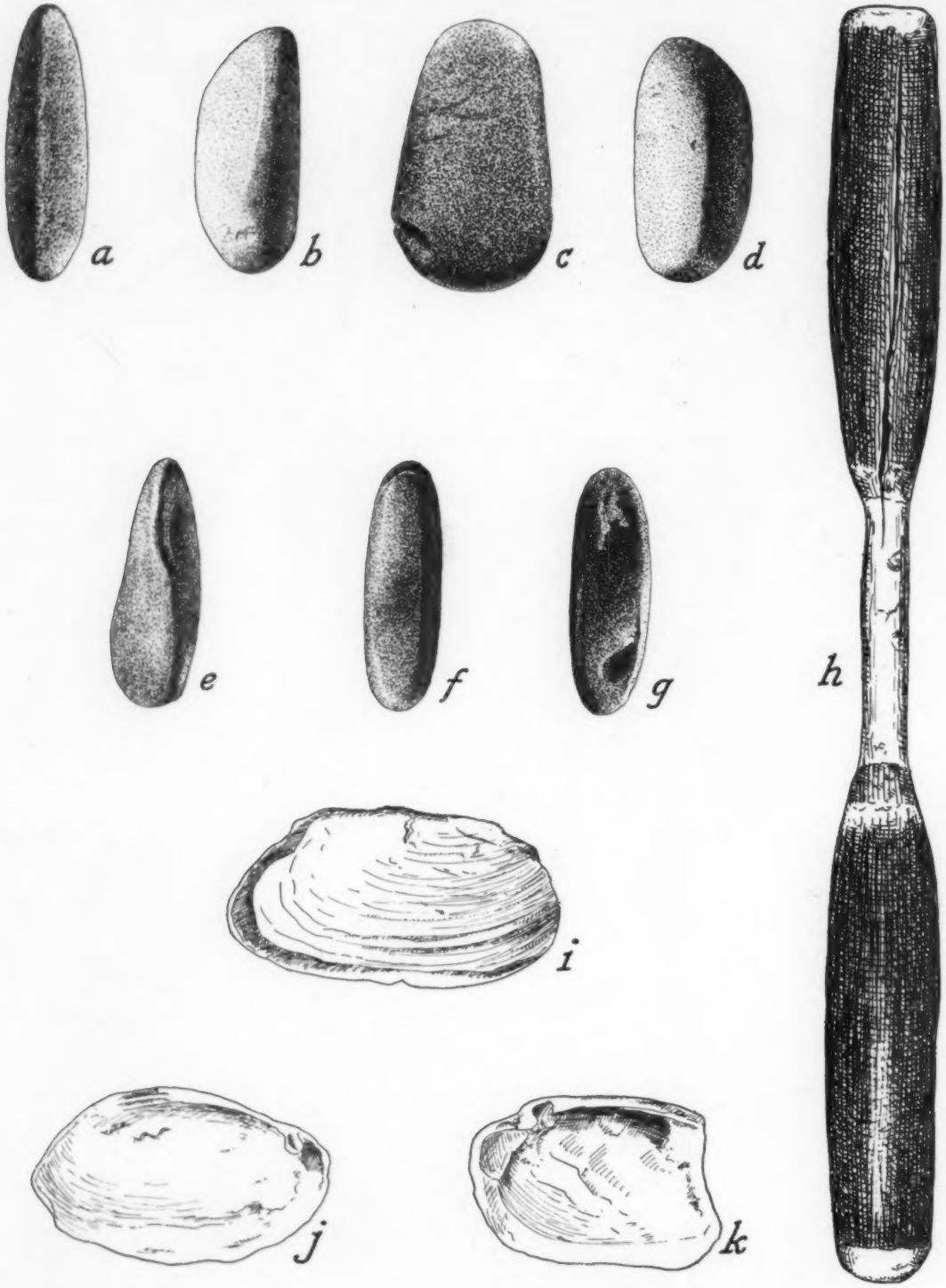


FIG. 1. *a–g*, polishing stones; *h*, wooden pestle for pulverizing clay; *i–k*, mussel-shell scrapers.

repeatedly spread and turned. The two types of clay are subjected to the pounding treatment individually.

For the manufacture of vessels a mixture of pan clay and pipe clay is used, the proportions being about half and half, or two thirds of pan and one third of pipe clay,[13] while pipe clay alone is used in the production of pipes.[14] The two clays are first combined by heavy pounding and by mixing; more mixing follows the addition of water which is applied after a satisfactory state of clay pounding has been reached. The raw material is not subjected to levigation,[15] nor are any solid substances added to the paste. In fact, the pan clay is arbitrarily freed of its natural sand by sifting and picking.[16] However, particles of mica, constituent to the clays, are allowed to remain; it would certainly be very difficult to remove them. Occasionally, blood of domestic animals is said to be added to the clay in mixing, after the customary amount of water has been applied; this, the natives hold, assures lightness of weight in the final product.[17] (Such blood is presumably highly coagulated and lumpy.)

Compacting of the paste is further advanced by hand kneading while the two clays are being mixed and water is being added. The preparation of the paste consisting exclusively of pipe clay undergoes a similar process. In either case the amount of admixed water is empirically governed by the degree of plasticity and consistency which the potter deems best for her purpose. There are no definite measurements, nor set standards, rather it is patent that the potter arrives at the desired state of satisfactory proportions more or less mechanically. Previous experience, quite obviously, is the determining guide. Upon the completion of the paste-producing procedure, the bulk is broken up into small lumps if manufacture of vessels follows immediately; otherwise it is stored away in larger quantities, usually bundled in cloth, to be drawn upon as needed. In the latter case, water may be occasionally added to replenish natural evaporation, but as a rule the paste is not stored long enough to necessitate this step. At any rate, old paste must always be re-kneaded to restore its pliability, and this inevitably requires additional water.

PROCESS OF MANUFACTURE

GENERAL PREPARATION

The potter operates either outside—generally on the open porch of the house, or under shade trees—or indoors, depending on weather conditions. Seasonal limitations, therefore, are rather insignificant. With the proper place selected, the necessary equipment and tools are assembled and placed within easy reach of the artisan, who works in a seated position. Of particular importance is a board used primarily for the preparation of fillets; this significant contrivance is here designated as the manipulating board. Some potters use yet another, smaller board, often called the "lap board," which is limited in function strictly to the purpose of supporting an embryonic piece in the building procedure, and in the shaping process. The essential tools comprise: a cane knife for cutting strips of paste and for general paring and trimming (fig. 2*g–h*); fresh-water mussel-shell (fig. 1*i–k*) and gourd-rind

[13] Harrington (1908*a*: 403) speaks of "about two parts of pan clay to one of pipe clay." *Cf.* also Holmes (1903: 53–54), where "about equal proportions" are recorded.

[14] Pipe clay alone is said not to be usable in the manufacture of larger pieces, for experience shows that specimens so produced "always" crack when exposed to fire. Apparently the natives have no remedy for this peculiarity, which seems to lie in the constituent properties of the pipe clay. However, the combination of the two types of clay affords an excellent medium with which to build (and fire) a vessel to satisfaction. The product of pipe clay alone is always a small specimen and its firing, in comparison with the average vessel, requires less heat. It is interesting to note that pan clay is never used alone. While this is not explained by the natives, it seems evident that its texture is recognized as unsuitable unless altered by an admixture of the pipe clay. Asked to operate with mere pan clay, Mrs. Edith Brown declined, saying that vessels so produced "would not fire to satisfaction."

[15] Even if the clay brought in from the pits is submerged in water while in storage, as is sometimes the case, the spontaneous "washing" which is started thereby is really negligible. So long as the potter's hands and tools are being moistened during the manufacture, a certain amount of levigated clay settles on the bottom of the receptacle. As a rule, however, this quantity is insufficient for practical application, and its fine quality is not considered an asset. In as much as the Catawba are not acquainted with the principle of the slip, this sediment is more or less wasted, unless it is eventually returned to the bulk of the stored raw material.

[16] In certain archaeological sherds collected on the reservation, I noticed the presence of calcined particles of crushed bone. As far as I know, such a medium has not been observed among the historic Catawba. Dr. John R. Swanton (personal communication) has recorded the use of burned bone (and also of crushed sherds) for inclusions among the Natchez.

[17] My attempts to secure a rational explanation for this reasoning from the potters were not successful. The practice was not demonstrated during my observations. The reference to the "lightness of weight" suggests that the admixture of blood clots adds an organic medium which in the firing process tends to increase the vessel's porosity.

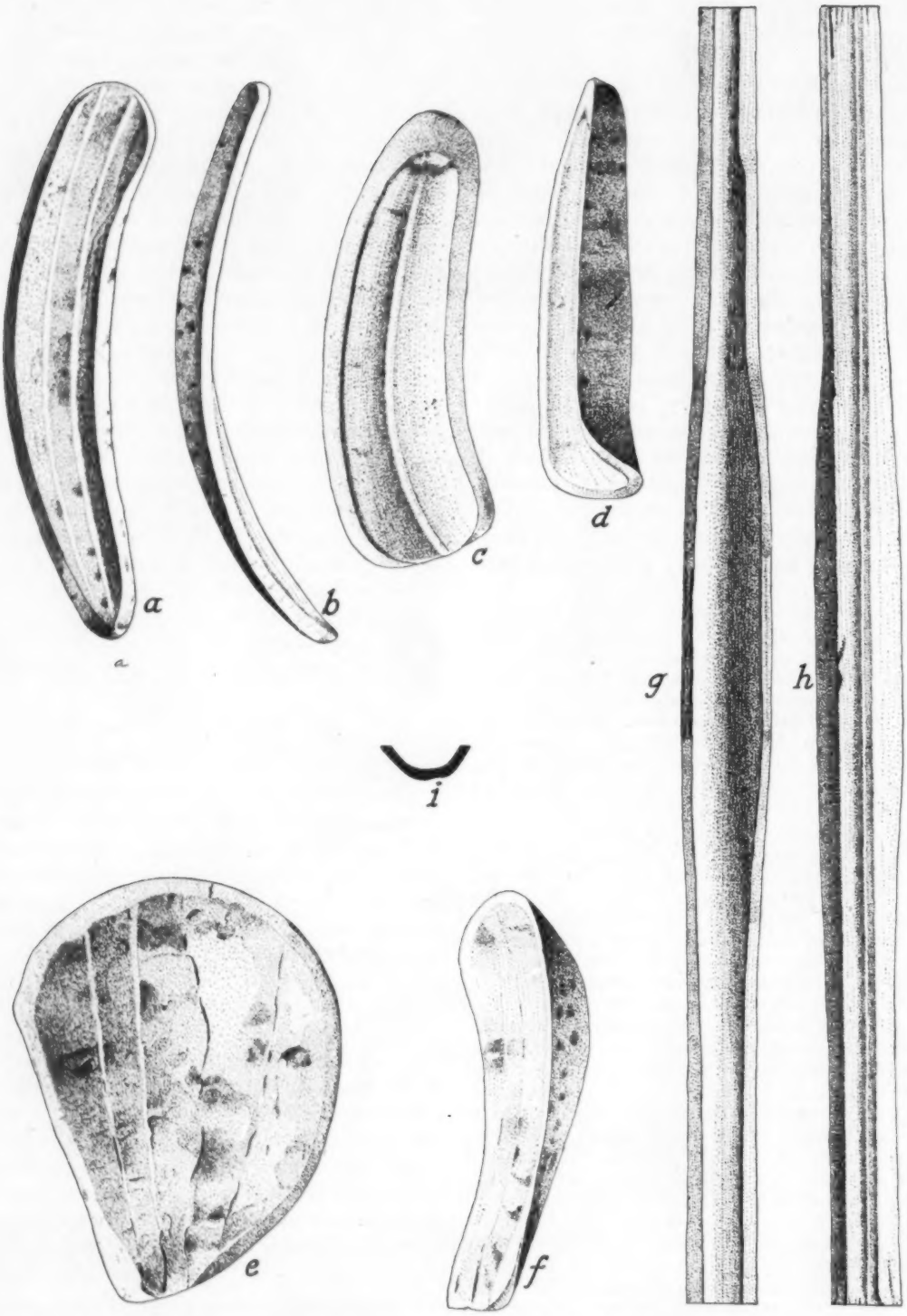


FIG. 2. *a–f*, three gourd-rind tools employed in form-giving; *g–h*, cane knife; *i*, cross-section of *g*.

scrapers (fig. 2*a–f*); polishing pebbles (fig. 1*a–g*).[18] A supply of water, usually in a metal receptacle such as a tin can or a wash basin, is constantly maintained at the potter's side for moistening the paste and tools, as well as for washing hands.[19]

BUILDING TECHNIQUE

It is not always possible to segregate pottery-building technique (*i. e.*, the actual process of construction) from shaping (*i. e.*, that manipulation which results in form-giving). With the Catawba the two tasks are sometimes accomplished wholly separately and sometimes, consciously and unconsciously, concurrently. For this reason it is necessary, at the outset of the description of the relevant processes, to make due allowance for the often inseparable duality involved.

The Catawba employ the following three categories of construction: (1) modeling, or direct form-drawing, out of a lump of paste, not necessitating subsequent shaping; (2) segmental building, depending on two annular variants, in both of which individual fillets of paste are used; (3) molding within a double form, used exclusively for the manufacture of certain pipes and lugs. Two distinctly different modes of fillet construction are commonly practised on the reservation: (1) the uninterrupted process, which uses the ring or the circuit variant; (2) the sectional process, which depends largely on the circuit variant.

The three categories and their respective variants were known to all the potters whose procedure I have observed. A detailed description follows.

Modeling

The modeling begins with a roughly spherical lump of paste, within which the potter first forms a grip depression to accommodate her thumb. The depression is worked into a cavity by pressing the fingers of one hand into the lump, and turning it with the other hand. Thereupon building and shaping progress simultaneously, both being the result of additional finger manipulation. The body of the vessel may be made either entirely in the potter's hands, or with some such support as a basal fragment of a broken vessel or a portion of gourd rind. It seems advisable to stress the purely non-form-giving function of such a support in order to preclude possible misunderstanding. The shape of the basal portion of the vessel under construction is not achieved by pressing the paste into the cavity of the supporting device. It is, of course, entirely possible to effect certain incidental shape conformity; however, the bottom has already been modeled when the support comes into use. The modeled specimens are usually of small size and simple in shape (figs. 3, 9).

The so-called "peace pipe," a Southeastern oddity,[20] consists of a globular bowl with four or more appended stem tubes (fig. 7*a*),[21] and imi-

[18] *Cf.* also Harrington (1908*a*: pl. XXIII*a*, *b*): shells; (*c–e*): shaping artifacts of gourd rind; (*f*): wooden implement, not explained by the author—perhaps a shaping or smoothing tool?; (*g*): iron knife, *i. e.*, one with a metal blade and a wooden handle; (*h*, *i*): cane implements, *i. e.*, both fashioned of splints; (*h*): apparently a knife; (*i*): perhaps a perforator?; (*j*): wooden perforator; (*k–m*): smoothing stones; (*q–r*): bone implements, evidently polishers, used in the manner shown on plate XX*k*. For polishing pebbles from Tidewater Virginia, *cf.* Speck (1925: fig. 106*a*, *b*): Mattaponi and (*c–g*): Pamunkey, shown in comparison with three Catawba specimens (*j–l*).

[19] Dr. Speck supplied the following interesting belief which he recorded at the Cherokee reservation from the late Mrs. W. West Long: "If a woman engaged in pottery making touches or handles dead mice or rats, pollution follows and causes serious breakage during the firing process. In order to alleviate such consequences the woman must wash her hands on four successive mornings in water procured from the holes of red crawfish, *i. e.*, pure deep water." The Catawba, as far as I am aware, do not subscribe to this belief. Yet, according to Mrs. Sally Gordon, although not corroborated or substantiated by others, dead mice or rats should be interpreted by the potter as "a bad omen, if anything."

[20] Originally also in use among the Chitimacha; *cf.* Swanton (1911: 349).—Whether or not its Catawba occurrence is to be related to the stone or pottery "peace pipe" mentioned but, in so far as the pottery variety at least is concerned, inadequately described, by Timberlake (1765: 39), cannot yet be established.—An interesting specimen of steatite, having four perforations for stems, drilled at equal distances from one another, is known from a grave find in Philadelphia; *cf.* Barber (1878: 113). This, as far as I understand the distribution, seems to be the most northerly appearance, yet known, of the type of pipe which accommodated four smokers at a time.—The Pamunkey probably borrowed the four-stemmed pottery variety from the Catawba; *cf.* Speck (1925: 427 ff.).—Speck (1925: 432) contends "the 'peace pipe' was a native Southeastern object." This view finds some support in certain scanty, yet suggestive archaeological evidence from the territory of the contemporary Catawba. Fragments of bowls and stems of such "peace pipes," collected during our stay at the reservation in ploughed fields, tend to endorse the deduction first advanced by Speck. However, the locality in question did not yield adequate evidence with which to establish, even provisionally, whether or not a definite archaeological site exists there. Moreover, the material may well be of modern Catawba manufacture, or in any case historic.

[21] For a six-stemmed, four-footed specimen, *cf.* Holmes (1903: pl. CXXVIII, lower row, center); the teatlike feet, of which three can be seen in the illustration, have the

FIG. 3. Construction of a modeled Catawba bowl from the initial lump of clay to the finished product. The stem and bowl of the pipe were carved out of a molded blank.

tates the shape of a spear head. This and sometimes the plain pipe (fig. 7*b*),[22] as well as miniature vessels, are produced by hand modeling. Similarly, bottoms of vessels, irrespective of size or form, are modeled in hands; they are first patted in the palms, then flattened into a discoidal shape, and finally neatly rounded (fig. 5).

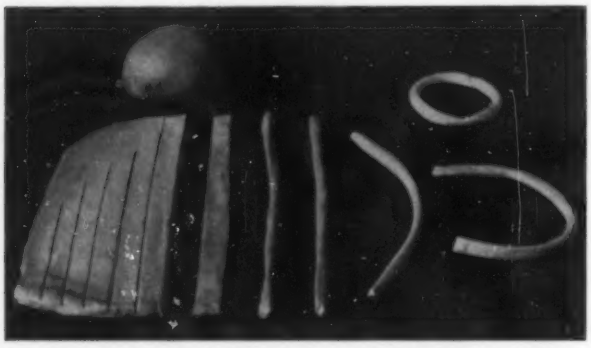


FIG. 4. Procedure followed in the fashioning of a single "ring" from the original lump of paste.

The technique of modeling complete vessels is a very simple process in which the fingers of the operator are the sole means of building and shaping the form, thinning the wall as it is being drawn out, and finishing the surfaces. Naturally, this requires rather scanty attention at the end of the shaping, for the potter's fingers have automatically smoothed the wall concurrently with its growth. There is also a minimum of scraping, if, indeed, any significant roughness does occur.

Only polishing, if desired, is finally done with the aid of specialized tools, such as a pebble, bone, or cane implement, or the specimen may be treated with a wet cloth to attain smoothness. Beginners, as a rule, are instructed first in the modeling method which is, quite naturally, considered the simplest.

Segmental Building

I have chosen this expression as a substitute for the *general* term "coiling." Its collective connotation includes *coiling proper* or that process of pottery-building in which the paste medium, either a single fillet or a series of successively joined fillets, linked as the potter proceeds, is wound spirally. (*Cf.* the section on "Coiling," *infra.*) It also includes the Catawba practices here called the ring and the circuit variants, either of which these potters employ in two distinctly different building progressions. The two modes of construction are here labeled, respectively, the uninterrupted process and the sectional process. Since a decision as to which of the two methods to be followed is necessarily determined by the potter in advance of the construction, it seems imperative to classify the fillet processes first of all in recognition of this principle. The headings of the description which now follows are, therefore, arranged accordingly.

THE UNINTERRUPTED PROCESS—

Strictly speaking, this process concerns only the erection of the wall, for irrespective of the variant followed by the potter (either ring or circuit), the bottom is always prepared first by hand modeling (fig. 5).

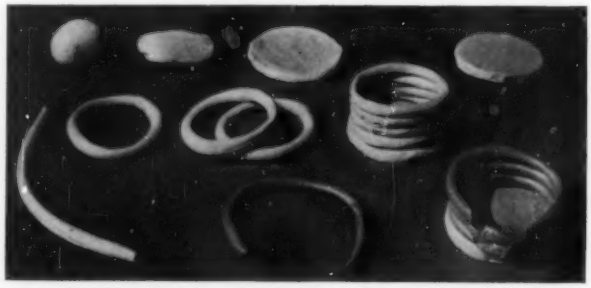


FIG. 5. Steps in the construction of vessels by both the ring and the circuit methods.

shape of a truncated cone; the four tubes in full view have a cylindrical form; the grooved decoration presents a linear design placed on the bowl (cross hatching, running chevron, and myrtle twig) and on the stems (four parallel grooves on each one of the visible four); a pronounced surface lustre is clearly in evidence. The specimen was collected either between 1876 and 1886 or later (Holmes, 1903: 143), *i. e.*, possibly, *by inference*, as late as 1903, the date of Holmes' publication.

[22] For illustrations of certain phases of pipe-making by modeling, *cf.* Harrington (1908*a*: pl. XXII*a–d*).

The ring variant.—The ring method consists of the superposition of several individual rings of more or less constant dimensions, each first separately fashioned and fully closed on the manipulating board. The construction of a prospective vessel begins with the modeling of the bottom, from a suitable piece of paste, in the potter's hands. The disk-shaped piece is placed on the supporting board with an impact which immediately flattens its underside. The board[23] usually has a coating of old, dry paste so that moistening to prevent adhesion of the disk is not always necessary. If the force of the impact causes distortion to the shape or to the margins of the disk, the potter carefully rectifies this and then proceeds to build the wall.

The preparation of the rings themselves involves the following manipulation. First, cylindrical fillets of paste, fairly uniform in length, are fashioned either beforehand, or as the process advances; as a rule, the manipulating board is utilized for this need. To prepare a fillet, a ribbon is cut, with a cane knife,[24] from a strip of paste which was first flattened out of a lump. The ribbons are then either rolled out on one of the boards, or manipulated between the palms while hanging pendant; thus the fillets acquire fairly uniform thickness and length, both of which are determined by the dimensions of the desired vessel. The fillets are closed into rings on the board prior to their entry into the building process proper; their diameters are more or less equal in each individual cylindrical blank. The ends of the fillet are somewhat thinned and flattened to prevent increase in the girth of the overlap, and any surplus paste is removed.

The initial ring is placed upon the previously flattened disk-base in such manner that it rests close to the horizontal edge thereof; the ring is then pressed down as soon as its outward margin coincides with the planed circumference of the embryo bottom. Subsequent rings are superimposed one by one, concentrically, and each is individually pressed down vertically, until the preconceived height is attained. The joints of the rings are usually well aligned in order to "aid in the shaping process which follows later"; at least so the potters explain this seemingly irrational peculiarity.

The potter achieves satisfactory bonding of each individual ring by pressing, pinching, and smoothing its surface with her fingers. Viewed in cross section, the traces of such bonding are parallel and reveal symmetrically overhanging spreads on either side of the underlying ring.[25] Either a slightly convex or a slightly concave curvature, or possibly an irregular flattening, is discernible in the profile of the zone marking the contact between the initial ring and the base. The overhanging then appears only on the outward side where the potter had pushed some of the ring's paste over the edge of the basal disk.

Throughout the bonding procedure, the artisan operates up and down the growing wall. Traces of the junctions are, of course, at least partially mutilated by the subsequent shaping and scraping processes. Nevertheless, a cross-section of the fired product often enables the observer to distinguish several criteria of the constructional technique. The ring joints, for example, are often visible through marks retained within the span of the overlap; individual rings are at times recognizable, and so are also their marginal overhangs (fig. 10). Sometimes a chip on a fired specimen exposes certain technical details. Dissection of a vessel still in the plastic state usually reveals ample evidence of the building process employed, despite the distortion occasioned by the scraping and form-shaping steps.

The circuit variant.—The circuit variant involves the use of strips of paste, rolled into fillets on the manipulating board, and fairly uniform in length. The following lengths of rolls have been observed: by Palmer, 12 inches or less;[26] by Mooney, 8–10 inches;[27] and by myself, 8–16 inches (0.203 m.–0.406 m.). It is to be remembered that unless sectional building is necessary, the potter usually produces first a cylindrical blank which is subsequently shaped into a final form. For the annular construction, therefore, the length of paste rolls, whether intended for rings or circuits, must be rather constant. Naturally, occasional deviations can be easily rectified.

The building proper again begins with the formation of the bottom. The first fillet is applied in a manner similar to the placement of the initial ring, but the completion of the circuit occurs on the growing embryo. A right-handed potter[28] proceeds in a clockwise direction. She holds the

[23] First recorded among the Catawba, as far as sources show, by Dr. E. Palmer in 1884; *cf.* Holmes (1903: 55).

[24] A metal knife is sometimes substituted.

[25] *Cf.* Holmes (1903: fig. 30*f*, *c*) for an analogous, schematic illustration.

[26] *Cf.* Holmes (1903: 55).

[27] *Cf.* Holmes (1903: 54).

[28] All the Catawba potters observed by me were right-handed.

fillet in the right hand, adjusts its placement and contact with the left, and also turns the bottom, this time counterclockwise, with the left hand. The fillet is somewhat flattened as it is pressed down and bonded. Its ends meet at the completion of the circuit, having been trimmed by finger pinching as the last step preceding their joining.[29] Excess paste is broken off; should there be a shortage, the potter either removes and re-shapes the fillet, or, if the deficiency is negligible, adds the necessary amount of material. Subsequent building proceeds by superposition of individual fillets; their placement, bonding, and alignment are accomplished in a similar manner as in the case of the rings. In reality the ring and circuit variants are functionally identical. In the one the ring is closed and then applied as a part of the wall-erecting process, whereas in the other this operation is reversed (figs. 5, 6).

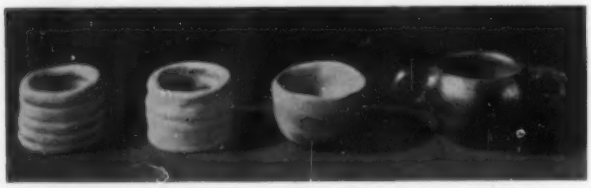


FIG. 6. The procedure followed in the construction of a Catawba vessel from the superimposed rings of paste at the left to the polished and fired bowl at the right.

During the building manipulation the hands of the potter are frequently washed in water, whereby they are not only kept free of paste, but also in condition to facilitate bonding and removal of surplus material. By the time the desired height is reached, the crude product, fully plastic, usually has a cylindrical form, distorted here and there owing to the frequent turns and patting by the operator. The fillets become planed on all sides soon after joining, yet they remain individually recognizable despite the frequently repeated hand smoothing. With the placement and bonding of the terminal ring or circuit, the annular process completes the erection of the wall.

Then follows shaping of the form.[30] Employing a spoon-shaped piece of gourd rind (fig. 2*e–f*) on the inside of the cylinder, and her free hand on the outside, the potter gently forces out a definite form, progressing, as a rule, upward from the bottom.[31] Superficial traces of the fillets are more or less completely obliterated by this operation. However, individual demarcations of their adhesions are retained inwardly; this is so because the pressure exerted by the potter upon each ring or circuit registers a lessened effect within the wall.[32] (Penetration of tools into the interior of the wall is entirely out of the question. During the placement and bonding of the individual rings or circuits only the fingers are employed.) Throughout the shaping process the tools as well as the hands are frequently submerged in water. In the removal of excess paste the potter utilizes a cane knife. When finished, the definitely formed vessel has a smooth appearance both inside and outside. The wall is now thinner, but there still remains a superfluous quantity of paste smeared over the body. The next step, therefore, involves a scraping process[33] whereby this excess is removed with the aid of a mussel shell,[34] which is also frequently dipped in water. As a rule the scraping is done on the outer surface first, because in the case of defects limited reshaping becomes necessary and its execution is effected by operating largely within the interior of the specimen. The scraping completes the shaping process, and at the same time prepares the vessel for its final surface finish, namely smoothing and polishing. The mussel shell also serves to cut and smooth the rim, for the lip form of which a provision is made as the final fillet is bonded and partially drawn out or perhaps everted. Vessels with a plain, undiverted rim,

[29] *Cf.* Holmes (1903: 54) (Mooney's observation): a roll "properly flattened out was carried around its [the disk's] circumference and broken off on completing the circuit"; and p. 55 (Palmer's observation): "One of these [rolls] is wrapped about the margin of the disk and worked down and welded with the fingers, and others are added in like manner until the walls rise to the desired height." Harrington (1908*a*: 403) speaks of "coiling" individual rolls one by one—in reality the circuit method.

[30] Although the description of this process should perhaps be reserved until the molding method has been presented, its inclusion at this point is motivated by a desire to retain continuity.

[31] The Catawba do not use a trowel or anvil, of either pottery or other material. As far as I am aware, there are no suggestions or positive evidence to show that either contrivance was known to their predecessors. The pottery trowel, as found archaeologically in the Middle Mississippi Valley area (Holmes, 1903: 35–36, figs. 6, 7 (diagrammatic depiction of a method of use), and pls. XXXV, XXXVI), is paralleled in Tennessee, *presumably* amid Cherokee remains; *cf.* Harrington (1922: 194, fig. 33).

[32] Tests on plastic as well as on fired (fresh and old) material clearly illustrate this point.

[33] This step, no longer a part of the building process, is inserted here for the retention of continuity.

[34] A piece of gourd rind, similar to the usual shaping tool, but with sharpened edges, is sometimes substituted.

or entirely without a lip, are terminated at the mouth by hand pressing or by pebble polishing. No drying is necessary before scraping begins, for this step has the function of reducing and smoothing the wall, as well as further increasing the total compactness. The state of complete plasticity, therefore, facilitates this treatment most advantageously.[35]

THE SECTIONAL PROCESS—

Vessels of large proportions and severe curvatures are produced by sectional building and subsequent joining. The component sections of such specimens, *i. e.*, the base with a lower portion of the belly, the shoulder with an upper portion of the belly, and the neck, are separately erected usually by the circuit, rarely by the ring, variant of building. These parts are individually shaped and scraped, and then joined. To effect their complete bonding, the potter works with a gourd rind tool wherever the hand fails to reach. It is thus possible to manufacture a great variety of body shapes.[36]

Molding within a Double Form

The mold is made of paste comprising equal parts of the two clays used in the manufacture of vessels. The molds are now rather rare on the reservation. Those still to be found were mostly inherited, and the younger people do not seem to be very skillful in producing new specimens.

Harrington (1908*a*: 405–406) recorded the following process of manufacture: "For making pipe molds an original model is shaped by hand, and after being burned in the usual way is greased and forced down into a flattened cake of fresh clay until half embedded, then the surface of the cake is also greased to prevent sticking, and another cake laid over and pressed down, forming a complete form of the original pipe. When dry these half molds are removed and burned, then they are ready for use."[37] Harrington is to be credited with the first record of the Catawba pipe mold manufacture. It may be added that some molds have pyrenoid protuberances fitting into corresponding depressions to aid in retaining the position of the tightly closed double form. Other molds have orifices bored through their corners to serve a similar purpose by the insertion of wooden sticks[38] (figs. 8, 11*e–f*).

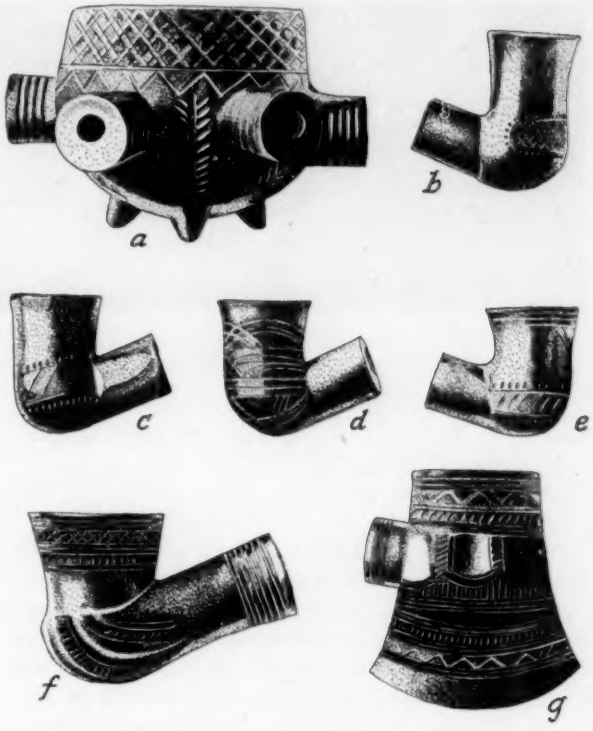


FIG. 7. Catawba pipes illustrating decorative styles: *a*, "peace" pipe; *b–f*, smoking pipes; *g*, pipe in form of axe.

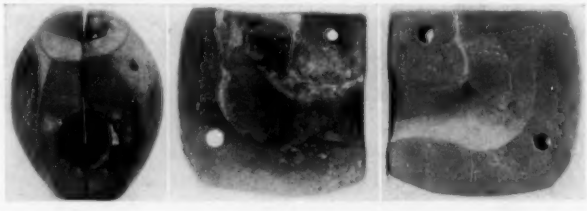


FIG. 8. Catawba pipe mold. (Courtesy Peabody Museum, Harvard University.)

[35] For illustration of various manufacturing steps employed by the Catawba in 1908, *cf.* Harrington (1908*a*: pl. XIX*c*), "roll shaping"; (*d*), "application of initial roll"; (*e*), "application of roll in later stage of building," *i. e.*, the circuit process, by Harrington, however, labeled as "coiling"; and (*f*), "blending of individual rolls," *i. e.*, superficial bonding of the fillet circuits; and (XX*g*), "rim shaping"; (*h*), "rim smoothing"; (*i*), "scraping of outer surface"—in the instance shown, done by a male operator; (*j*), "rubbing with a pebble"; (*k*), "rubbing with a bone implement."

[36] The sectional process has not been recorded in hitherto published sources on Catawba pottery.

[37] For illustrations of pipe molds, *cf.* Harrington (1908*a*: pl. XXIII*n*, *o*, *p*): single half and two halves of the same mold—Indian head form.

[38] *Cf.* also Harrington (1908*a*: pl. XXIII*o*, *p*).

FIG. 9. Various types of smaller Catawba pottery pieces, constructed for the most part by modeling.

The molds are now frequently borrowed from house to house. All varieties of pipe shapes, with the exception of the "peace" and the spearhead types, are usually molded, although the plain form is frequently hand modeled. The actual technique of molding may be described as follows. A rough shape, approximating the cavity of the form, is modeled by hand, and pressed into the mold in which it is enclosed as much as the bulk will allow. The excess paste oozes out between the incompletely closed halves of the form, is removed, and the pressure is renewed. The process is repeated until the mold is tightly closed. In the meantime the paste blank is frequently extricated and scraped, and inserted again until it assumes the final conforming shape. When satisfactory form is attained, the specimen requires only drying to be ready for surface treatment, for the piercing of the stem, and for the carving out of the tobacco bowl (unless it is to be used as a lug, which is often the case with the Indian head type) (figs. 13, 14).

Harrington (1908*a*: 405) speaks of coating the cavities of a pipe mold with grease or ashes "to prevent sticking"; such procedure was not noted

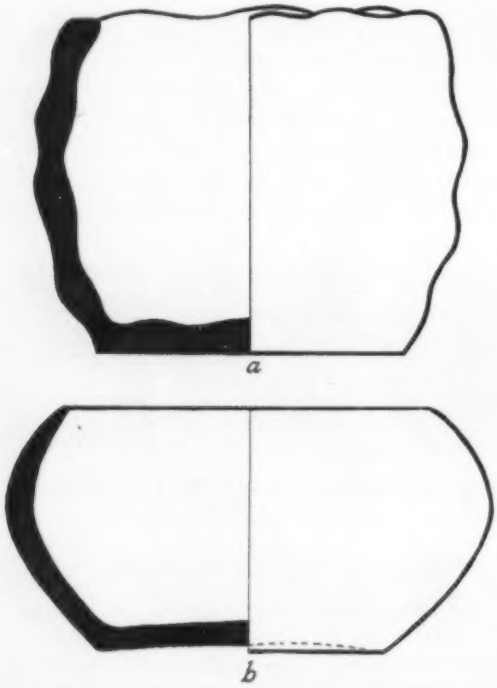


FIG. 10. Cross-sections of Catawba pots, indicating wall thickness and contour.

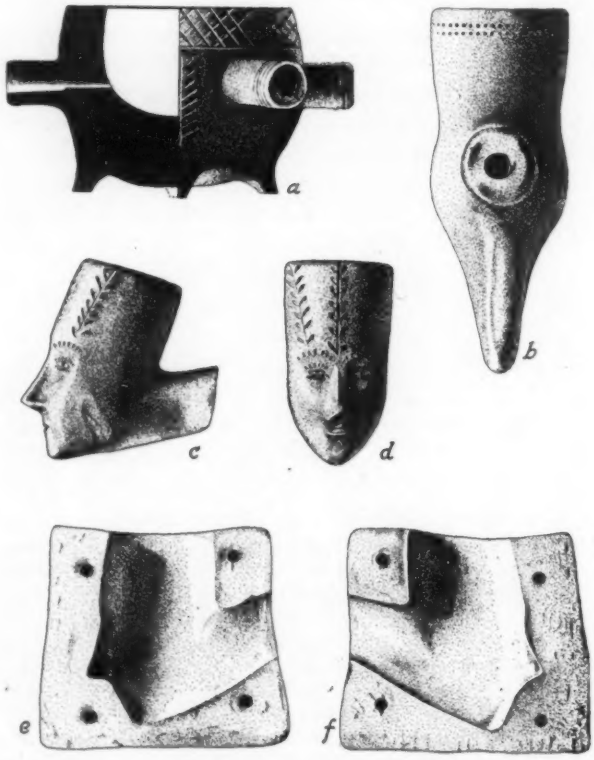


FIG. 11. *a*, "peace" pipe in section; *b*, pipe in form of spear; *c–d*, pipe in form of Indian head with stylized feather bonnet; *e–f*, pipe mold.

FIG. 12. Vases exemplifying the mottled effects achieved by the Catawba in firing: *a*, plain handles; *b*, unadorned cylindrical vase; *c*, flared lip; *d*, lugs in the shape of Indian heads.

by me, yet the potters did not seem to experience any difficulties in the alternately repeated insertion and removal of the plastic blank.

FORMS

The pottery forms of the contemporary Catawba seem to be determined by local as well as by external factors; the former are governed by tradition and the latter by economic forces. The bulk of the product is absorbed through commercial channels and is sold to the ultimate user chiefly through an intermediary. The market and its demand unquestionably exert certain influences on the form, and have done so, apparently, for some fifty or sixty years.[39] Thus candlesticks, book-ends, ashtrays, and various grotesque forms now so commonly produced, are concretely illustrative of this force. Nevertheless, purely native strains still dominate the craft. After all, the market is interested in this ware because it is Indian. The basically old globular and semiglobular forms are retained, protean tendencies are nominal, and above all the technique of manufacture remains wholly aboriginal (fig. 12). The guiding criteria which enable the student of Catawba culture history to affiliate their ancient ware with the modern product rest first of all upon the phenomena of technique. Although archaic forms are giving way to shapes dictated by the marketability of the product, the continuity of aboriginal methods is paramount. The peculiar process of fillet-building by the ring method appears to be decidedly endemic. However perverted the forms may now appear, their manufacture plainly documents an unmistakable uniformity in qualitative properties, building manipulation, surface treatment, and firing, all of which preserve elements recognizable in early historic Catawba pottery.

The contemporary forms may be classified into two groups:

(1) Endemically inspired similes of older examples, governed by traditional survivals,[40] in which globular and broad-bodied open jars, semiglobular bowls, conical bowls of low height, casserolelike vessels, each with a flat bottom, are

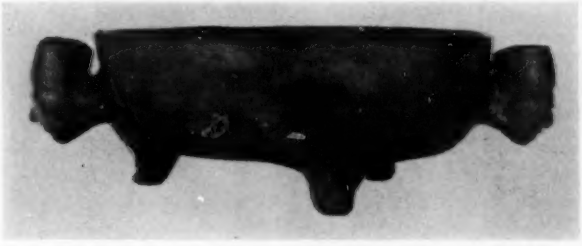


FIG. 13. Catawba vessel with plain lip and handles in the form of unadorned Indian heads. (Courtesy of Peabody Museum, Harvard University.)

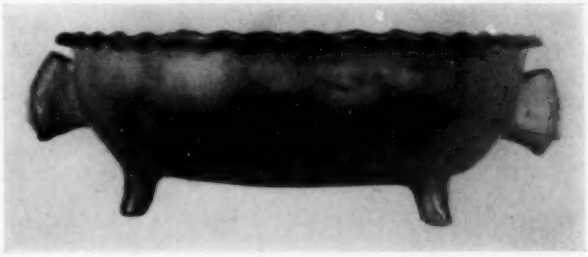


FIG. 14. Catawba vessel with flared indented lip and Indian head handles depicting feathered bonnet. (Courtesy of Peabody Museum, Harvard University.)

[39] *Cf.* Holmes (1903: 143); Speck (1934: 70).

[40] And up to the late eighties, apparently, perhaps also by actually retained ancient specimens, such as were then collected for the U. S. National Museum; *cf.* Holmes (1903: 143).

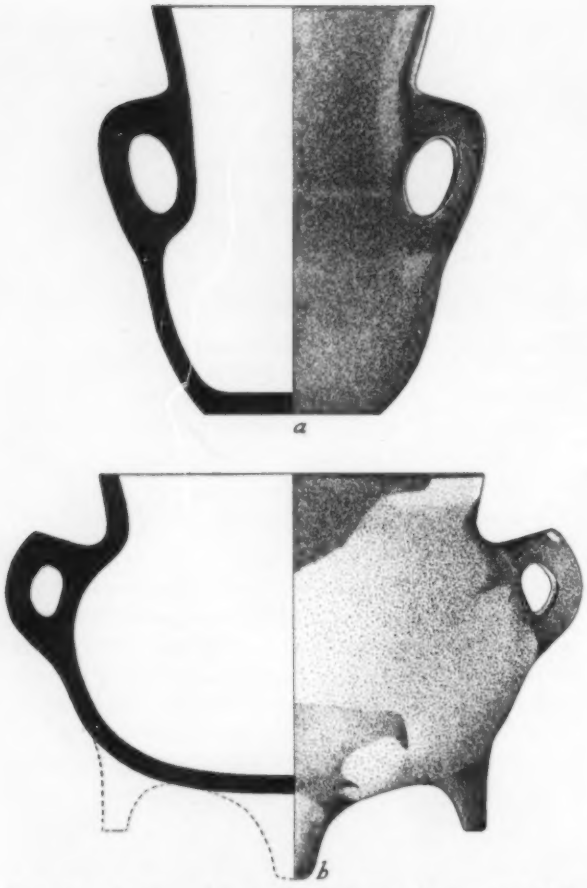


FIG. 15. Cross-sections of Catawba pots indicating thickness of walls.

common (figs. 11–17).[41] The jar sometimes has three or four teatlike pedestals, in which case the bottom is usually rounded (figs. 11, 13, 14).[42] There is no dependable archaeological evidence with which to consider the likely antiquity of this form. The drum possibility is perhaps to be held in abeyance; on the other hand, the shape may well have been derived from the European metal kettle or cauldron.

[41] The flat bottom seems to be an old characteristic in Catawba pottery; *cf.* Holmes (1903: 143). However, Harrington (1908*a*: 407) considered the flat bottom a sign of the modern product only, a view which, I think, requires revision. The flat bottom is characteristic of all pottery reasonably ascribable to pre-contact or early historic Catawba.

[42] For illustrations of this type, *cf.* Holmes (1903: pl. CXXVII*b*, rear row, first from the left), *i. e.*, a modern "Cherokee" piece (*vide* also *ibid.*: fig. 4, depicting a three-footed vessel converted into a drum); and further Harrington (1908*a*: pl. XXIII*c*). For Pamunkey similarities, *cf.* Holmes (1903: pl. CXXXVI, 1 and 3, counting from the left) and Speck (1925: fig. 114).

(2) Forms inspired or demanded by commercial opportunities, among which practically any shape may now be found (figs. 7, 9–11, 13–15). Again, the flat bottom is typically constant, and scalloped rims seem to be favored. The canoe shape may possibly be a modified survival of an elongated bowl; however, it may equally well have been developed in post-European contacts. Such pieces as ashtrays, book ends (figs. 20, 21), wall vases, and various other non-native forms must have been adopted by the Catawba, either directly or indirectly, since colonial days. Whether or not the pottery terminal of the syringe-insert[43] is aboriginal, is open to question.

The tobacco pipes include the most commonly made form with a plain bowl, the rim of which may, at times, be drawn out into a mildly flaring lip (fig. 22*a*); the rooster-comb shape (fig. 23*a*), in which the appendage may be either pronounced or moderate; the boot type (fig. 23*d*); the Indian head with a stylized depiction of a feather bonnet

FIG. 16. Catawba vessel in old native inspired form.

[43] Speck (1934: 51, note 1).

FIG. 17. Catawba vessel in old native inspired form.

(fig. 11*c*, *d*); the spear (fig. 23*c*); the axe (fig. 22*c*); and the so-called "peace pipe" (fig. 7*a*).[44] Of these various types of pipes, only the plain pipe has close form analogies among archaeological finds in the Southeast.

Such vessels as are still in active use on the reservation perform the function of general receptacles, storage facilities, and, in a decidedly limited degree, culinary utensils. In 1929 only Mrs. Sally Gordon preferred pottery vessels (of her own make) to metal receptacles. She found them, as she explained, "more satisfactory in general, and much easier to keep clean." The usefulness of native vessels as catch-alls was amply demonstrated in all the dwellings which I visited; both damaged and intact vessels were employed for such purposes.

Of the pipes, the plain type, in the short tube of which a stem of cane is inserted, is generally smoked.

According to Speck (1939: 50), "the older [Catawba] informants . . . all remembered something definitely of the pot water-drum . . . constructed of an ordinary clay pot of the Catawba ware of medium size over the mouth of which a piece of wet rawhide . . . was stretched and bound below the rim of the pot by a thong or cord wrapping."[45]

"In regard to the existence of burial mounds as recalled in the tradition of the oldest [Catawba] Indians questioned," writes Speck (1939: 45), "it may be noted that nothing has been brought to light. And the interment of pottery with the deceased is known to the present Indians only through pottery jars, of the types resembling those which they still produce, being found occasionally in graves washed out or invaded in the reservation cemetery." This observation is to be kept in abeyance for future archaeological work in the Catawba area.

Catawba pots vary in size from miniature vessels to pieces with a capacity of several liters. Wall thickness depends on form and dimensions; on the whole, 0.01 m. represents a fair average. (Book ends are often 0.02 m. to 0.03 m. thick.) Open bowls measure as much as 0.5 m. in mouth diameter, while their bottom diameter equals about two-thirds or less of that measurement.

FIG. 18. A particularly handsome three-legged Catawba pot.

[44] Illustrations of modern Catawba pottery will be found in Holmes (1903: pl. CXXVII*a*), showing "vessels collected betwen the years 1876 and 1886" (*ibid.*: 143), of which only the pitcher (rear, center) is not a representative of the otherwise truly native forms, and (pl. CXXVIII) pipes "of the same or a later period" (*ibid.*) including plain shapes and the "peace pipe," which is thus dated as being manufactured at Catawba prior to 1903 (*i. e.*, the publication date of Holmes' report). Also in Harrington (1908*a*: pl. XXIII (lower) *a–j*), of which only *b* and *c* may be considered as being due to European inspirations, while *a* and *d–j* represent the common native shapes. The boot-shaped pipe from Pamunkey, illustrated by Speck (1925: fig. 123, lower, extreme right), has a typical Catawba form, and was likely either made by a Catawba potter or copied after a Catawba sample; *vide* Speck (1925: 414 ff.). Speck (1925: fig. 125*d*) shows a Catawba pipe of the Indian head type, apparently a molded specimen, in comparison with Pamunkey imitations (*e* and *f*), both produced by hand modeling.

[45] *Cf.* also Speck's figures 1 and 2, illustrating improvised pot drums.

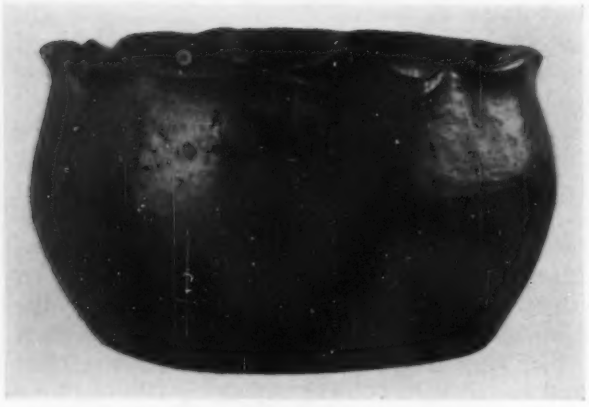


FIG. 19. Unadorned Catawba vessel with everted, flared rim.

Many jars reach a height of 0.4 m.; a very few, even 0.5 m. or 0.6 m. In Speck's *Texts* (1934: 71) a traditional recollection of "large milk pots" is recorded; these are no longer made, and their size remains conjectural. As far as I know, archaeological material of presumed Catawba manufacture does not reveal dependable data with respect to the sizes of the ancient ware.

SURFACE FINISH

The final surface treatment, executed while the specimen is still in its plastic state, consists of smoothing with a moist rag and polishing with a cloth or with a pebble[45] (fig. 25), unless decorating takes place. However, the Catawba usually omit embellishment of vessels.

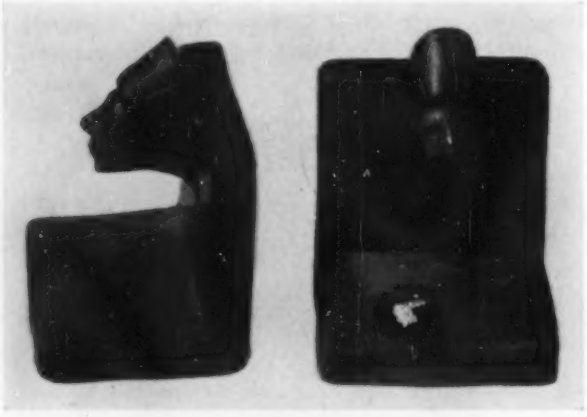


FIG. 20. Catawba book-ends.

[45] This involves mechanical pressure, and the resulting friction always leads to accelerated compacting of the surface so treated, which increases luminacy and, upon firing, leads to a lustrous effect.—The rag is of coarse material, such as burlap, while the cloth is invariably a piece of some soft cotton goods.

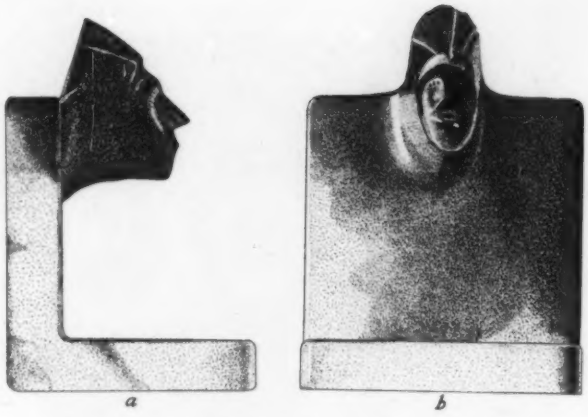


FIG. 21. Book-ends with Indian heads in relief.

The rag smoothing is usually applied to the outer surface alone. A piece of old sack is commonly used for the purpose; excess paste is removed, cracks and defects are obliterated, and evenness of wall is attained. The chief function is to prepare the piece for the polish.

The polishing is executed over a moistened surface with the aid of a worn, smooth pebble

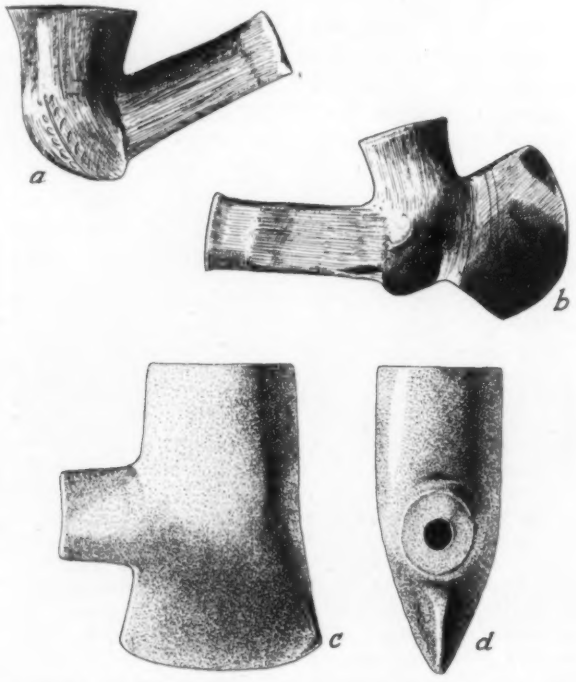


FIG. 22. Catawba pipes: *a*, plain with slightly flaring lip; *b*, molded in the war-club shape; *c–d*, in the shape of an axe.

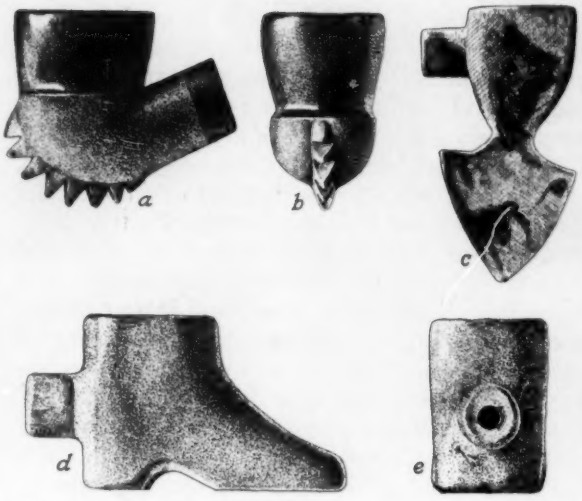


FIG. 23. Catawba pipes, molded in the form of a rooster's comb (*a–b*), a spear (*c*), and a boot (*d–e*).

(figs. 1*a–g*, 25, 26), a bone tool made of a rib or of a shank splinter (fig. 26), or a stick of hardwood stripped of its bark.[47] A piece of cloth, or of a soft skin, vigorously applied over a thoroughly moistened surface, will produce a polish comparable, in compactness and lustre, to one attained by the tools just stipulated. However, the cloth and the skin leave very inconspicuous identification stroke marks as against the typical pronounced facets of the stone, bone, or wood tools.

The polishing pebbles are regarded as somewhat of a precious possession and are retained in

FIG. 24. Catawba potter employing ordinary kitchen knife as scraper.

[47] The metal handle of a table knife is sometimes substituted with equally effective results.—*Cf.* Harrington (1908*a*: pl. XXIII*q*, *r*, and pl. XXII, 1) for illustrations of bone polishers; and his pl. XXIII*k–m*, for illustrations of polishing pebbles.

FIG. 25. Catawba potter using a polishing stone. Note bone scraper lying beneath right hand.

a family often for several generations.[48] They are ovoid, elongated quartzitic pebbles, blunt at either end, and with one or more working facets. The direction of the smoothing stroke depends upon the size and shape of the vessel under treatment; vertical and horizontal movements are most common. Generally the process is restricted to the outside surface and to a limited zone below the rim on the inside. Viewed in cross-section immediately after execution, the smoothed margin stands out in distinct contrast to the rest of the wall. A smoothed area is distinguishable through an accelerated textural density in the paste so affected. A similar difference is likewise discernible in the cross-section of the fired product. The textural difference, of course, is then especially prominently brought out by the heat. In extreme cases this phenomenon simulates a slip, which, in the true sense, is completely unknown in contemporary Catawba pottery, and, as far as can be ascertained at the present, is also absent in the ancient ware.[49]

[48] *Cf.* the statement in Holmes (1903: 55) that Catawba women married and living at Cherokee had brought smoothing pebbles with them from their homes, and (p. 56) that such a stone was in use by a Cherokee potter's family for three generations. *Cf.* also Du Pratz (1758, 2: 179) regarding the careful preservation of smoothing pebbles among the potters of Louisiana. For a Pueblo analogy, *cf.* Guthe (1925: 28); for South America, Linné (1925: 107). Smoothing pebbles, with unmistakable facets documenting their original use, are known from sites around the Catawba reservation.

[49] In old Catawba pottery, polishing frequently appears on the interior which, in contrast to modern pieces, is often quite dark. The possibility is to be considered that the Catawba originally may have "smoked" the interior of

FIG. 26. Catawba potter smoothing interior of vessel just below the lip. Bone scraper and polishing stone in foreground.

Pebble polishing is a particularly constant feature in the treatment of pipes, which are given a highly lustrous finish; temporarily "lost" in firing, *i. e.*, subdued by a carbon film, such lustre is easily reinstated by wiping the cool specimen with a dry cloth.

DECORATION

While Catawba vessels are predominantly unembellished, smoking pipes are very often decorated by rather simple incising, grooving, fluting, and impressing. The patterns are geometrical, combining rectilinear and to a lesser degree also curvilinear lines in the case of the incised, grooved, and fluted techniques. The impressed variety is attained by rocking or rolling a corn cob over the surface. Another effect is sometimes produced by rubbing the vessel with a corn cob which is firmly gripped in the operator's hand. While modern means, such as the milled edge of a coin, a hair pin or a shoe-button hook, a piece of wire, etc., are frequently called upon, native tools, *e. g.*, the mussel shell, gourd rind, and cane knife, nevertheless preserve their importance. In all cases the embellishment is executed while the specimen is still in a plastic state, after the polishing, and usually upon a moistened surface[50] (figs. 7, 11, 22, 23).

Among the designs given the pipes, floral motifs, expressing single or multiple petals, leaves, etc.; curvilinear combinations, such as cartouche, arches, undulating lines, etc.; and rectilinear figures, forming herringbone or myrtle twig simulations, zig-zags, chevrons, diamonds, rectangles, etc., are common. Single lines or multiple parallel lines are frequent, and bands defined by two lines in which the inner space is filled with cuts placed athwart seem to be especially preferred.[51] The technique of decorative execution in any one of the several varieties of design just enumerated employs incising, fluting, grooving, and impressing. Frequently the incised work is so delicate that it approaches engraving. The impressing technique depends chiefly on the looped end of a hair pin. Juxtaposition of differently produced decoration on the same specimen is by no means rare. The tools used in the decoration of pots and of pipes are about the same; the edge of a sharp knife, a piece of wire, and a hair pin, however, are particularly preferred in executing the more delicate designs on pipes.

Decoration of those pipes which require hand or tool shaping of their outward features in applied relief is usually deferred until the details are finished. (*Vide infra*, under drying process.) On the whole, however, such pieces remain unembellished unless they receive some such simple additions as parallel lines, rows of short cuts, serration of the lip or of the rim, etc.

The striking paucity of decoration on vessels is minimized, and to a degree compensated, by the colorful final product which acquires a vivid mottled effect in firing, and is often intensified by a high degree of lustre. The mottling is controllable and appears to afford much satisfaction to the aesthetic desire of the potter, who really values it as a decorative asset. The lustrous quality serves to enhance the mottling.

DRYING PROCESS

With the surface finish (and decoration) completed, the specimen is set aside for drying. The length of time required for this purpose varies from one to several days. In certain instances, large vessels may remain in drying a full week.

vessels in a manner perhaps comparable to that described for the modern Cherokee by Harrington (1908*b*: 226); *i. e.*, by smudging the interior with a slow fire of corn cobs.

[50] Harrington (1908*a*: pl. XX*e*) shows a photograph of a potter engaged in decorating a vessel which, however, reveals no details of the design nor the technique involved; Harrington explains (*ibid.*: 404) that the edge of a cane knife was actually used. His illustration of a pipe-decorating process (*ibid.*: pl. XXII*d*) likewise fails in details.

[51] For sample illustrations of decorated pipes, *cf.* Holmes (1903: pl. CXXVIII); specimens collected after 1876, *cf. ibid.*: 143.

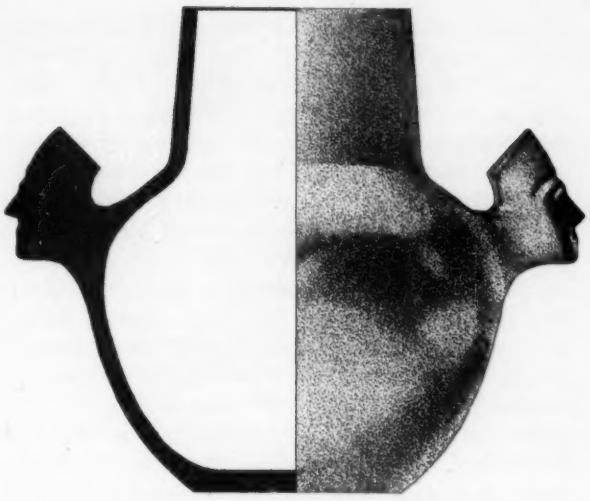


FIG. 27. Cross-section of Catawba vase indicating thickness of wall.

A shady, protected place outdoors, or a dry ventilated room inside, are preferred. Care must be taken that the pieces are not exposed to frost, which would, according to practical experience, inevitably cause chipping and fracturing in firing.

During the drying period, or sometimes after its culmination, the appending of lugs, handles, spouts, and legs takes place. The wall of the vessel selected for such additions is pierced in appropriate locations with the point of a cane knife or an awl. The appendages, comprising lugs and handles, are prepared by modeling or by molding. Their terminal tabs, ovoid, circular, or oblong in cross section, are inserted, secured and fastened from within, and the contact is carefully smoothed in and out to obliterate roughness. (Within the thickness of the wall, however, the bonding is less complete and leaves recognizable traces of the operation.) Sometimes these additions receive their final form before attachment, while on other occasions they are shaped by hand modeling after the appending is completed. Upon breakage, the fired pieces invariably reveal the shape of the tab and the manner of attachment when the retention within the wall is exposed.

The typical forms of the several categories of appendages are characterized by a rather small range of variety. The lugs, distinguished by a single plane of attachment, are predominantly conical and invariably blunt at the apex. The Indian head type, shaped in a pipe mold, is used as a lug (figs. 12*d*, 13, 14, 27), rather rarely.

The handles, always attached at both of their termini, are of two varieties: the band handle, having a strap-shaped body of an elongated rectangle in cross-section; and the rod handle, shaped like a cylinder or a rod, and circular, oval, or rectangular in cross-section. In either case, the handle usually describes a sufficient loop to accommodate a firm hand grip. A "wish bone" type of rod handle has been developed in the past few years; it appears on the slender, pedestaled vase (figs. 24, 25). The spouts range from a mere dent in the lip to a carefully modeled elaborate feature with a pronounced protrusion. The legs, on the whole, closely resemble the forms of the plain lugs; they are mostly conical, although at times cork-shaped, and invariably flattened (figs. 15*b*, 18).

Pipe blanks must be dry before the tobacco bowl is carved out and the stem is perforated for smoke passage, for otherwise the incompletely evaporated paste would be less resistant to distortion; its consistency, when thoroughly dry, retains a satisfactory degree of softness for the carving and piercing. The details of the relieved features, such as the chin, mouth, nose, eyes, brows, and halo of the Indian head type, the serration of the rooster comb, etc., are fashioned

FIG. 28. Mrs. Eliza Gordon (Catawba) at work on a vessel held on her lap. The scraping tool is a kitchen knife.

FIG. 29. Two young Catawba potters employing polishing stones.

by paring, scraping, and carving (fig. 28). (The paste so removed is used again in mixing fresh clays.) Positive appliqué may also be executed; thus, for example, the rooster pipe, which originally had its blank crest prepared in a mold, is now completed, by additional modeling and carving, into a rooster-comb effect. The potter does not find it cumbersome to alter, by addition or reduction, the original shape of a given blank. New or additional decoration is then executed as desired. The final manipulation preceding the firing of pipes involves cloth, bone, or pebble polishing.

All specimens undergo a careful scrutiny before firing begins. Flaws and defects are rectified, here and there a few strokes of smoothing and polishing are added, the security of appendages is tested, and meticulous examination is made of the state of dryness. While the potters do not seem to be conscious of any criterion with which to judge adequate dryness, it appears that they find guidance in the surface color. When a uniform tone of greenish-gray is shown, the specimen is considered ready for firing. Satisfied that the raw product has received ample care, the potter proceeds with preparations for the firing process.

FIRING

Ceramic pyrogenation of the Catawba depends on an open fire in the case of pots, and on a fuel-smothered process confined to a receptacle in the case of pipes. Even the latter mode, however, does not involve the principle of a kiln, and its temperature maximum is below that of an open fire.[52]

Some of the older potters still use a plain hearth, or rarely, a pit, out of doors; in any case, the method can be demonstrated upon request.[53]

While the domestic fireplace has now largely replaced outdoor firing, the technique involved is essentially the same, *i. e.*, it is still based on the principle of the open fire. The main difference between the two means seems to lie in the quantity of specimens which may be subjected to the heat at the same time; naturally, the domestic fireplace can accommodate a smaller number of pieces because of spatial limitations. On the other hand, the shelter of a dwelling is of signal advantage in the control of the fire, and facilitates operation irrespective of weather conditions. It seems to be generally understood by the potters that firing of vessels is not to be done on an excessively hot day; the reason for this is perhaps to be attributed to the physical comfort of the operator rather than to other considerations.

The fuel consists of wood, tree bark,[54] and corncobs. The firing process begins with the building of a substantial fire in order to establish a fresh bed of ashes which is localized more or less in the center of the hearth. The two classes of the product, pots and pipes, are always fired separately. With the fire well under way, the pots are assembled near by, within a distance of 1 to 1.5 m. There they remain exposed to the radiating heat for perhaps a half hour. Meanwhile the periphery of the hearth has been swept and the pieces have been advanced closer to the fire, being occasionally turned as well. They are, therefore, not subjected to extreme heat until a certain amount of slow evaporation has taken place. The vessels are eventually set close to the fire, in either upright, inverted, or reclining

[52] At any rate, with the Catawba ware, it seems proper to speak of firing rather than "baking." But it would certainly be wrong to say that these people (or any potters for that matter) "burn" their pots, for clay and inorganic inclusions are not normally pyrognomic.

[53] Harrington (1908*a*: 405–406) describes an outdoor firing operation specifically arranged for his benefit; similar procedure applies indoors. As a rule, vessels are not placed on the ash bed of an indoor hearth; such practice seems to go with firing outdoors (*ibid.*, and Speck, 1934: 70). For illustrations of an outdoor process, *cf.* Harrington (1908*a*: pl. XXI*m*), "preliminary heating of vessels"; (*n*), "vessels inverted upon embers"; and (*o*), "firing of vessels."

[54] Apparently preferred by some of the older potters; *cf.* Speck (1934: 70).

position, in the last-named instance with the mouth turned toward the flames. As an additional supply of fuel is being consumed, care is taken to confine the limits of the fire to the original bed. Yet, no effort is made to prevent an occasional leaping of flames over the vessels. Frequently the brisk fire expels sparks or bits of embers which settle and are allowed to remain upon the specimens. The contact of such hot particles produces a lasting mark, namely, a lighter discoloration of the surface, which remains recognizable in the finished product. The size and penetration of the spot so occasioned depends on the intensity of the heat as well as on the interval during which the necessary fuel medium remains in contact with the surface of the vessel. Sometimes this medium is applied intentionally, for the potter is well aware of the cause and effect involved. In the advanced stage of firing the hot pots are set practically next to the fire, being turned frequently in the meantime to gain even distribution of the heat. At no time, however, are the vessels placed within the fireplace proper where they would be most difficult to handle. The shifting and turning of the vessels is accomplished with a stick or with the rake which is used in attending the fire.

On two occasions Mrs. Sally Gordon smeared an incompletely fired, hot vessel with a grease-soaked rag, having first cooled it by removal to a comfortable distance from the fire. Asked for an explanation, she responded that such treatment produced blacker and lighter-weight ware. This was later corroborated by the qualities of the vessels in question: they were quite dark, yet mottled, and of a relatively light weight.[55]

Soon after exposure to the heat, the color of the specimens changes from greenish gray (dry state) to a reddish brown or grayish buff, and these tones darken as the firing, a matter of one to several hours, continues. These colors are the basic values, for the dark gray to black effect, which is so often attained, is really produced through a distinctly separate agency. As the specimens are advanced practically next to the fire, a process of carbonization is invoked as soon as fresh fuel, this time largely bark or wood chips, is added to the flames. This process produces much smoke and creates a reducing atmosphere. If the specimens are no longer turned during this procedure, only the sides exposed to the fire, and their margins, will be colored dark gray to black. The tone may be intensified by placing bits of bark directly upon the specimens with hot clinkers superimposed thereon. Occasionally the burning particle resting upon a vessel (or in close contact with it) generates excessive heat owing to a draft and increased oxidation, and promotes alteration in the surface color. In such cases reduction is prevented. Hence an area so affected becomes extremely light in color. In instances of this kind (limited in intensity and definitely allocated to small spots), color changes to light shades are often apparent in the course of the firing. More often, however, the burning particle of glowing fuel undergoes slow fire consumption, which produces sufficient gas and carbon to create the characteristic surface blackening.

In the case of pipes and smaller vessels, the entire specimen (at times a number of them at once) is placed within a receptacle, such as a metal pail, filled with bits of bark or wood chips; this fuel is then set afire, and allowed to burn for a certain period; unoxidized carbon is thereby retained. At times the burning contents of the receptacle are subsequently emptied upon an ash bed, and heaped over with additional, similar fuel. There the firing continues, again under a somewhat smothered condition and limited oxidation, and eventually results in fairly uniform blackening of the specimens. The pieces first smothered in the fuel within the receptacle undergo a process of slow firing under reducing conditions whereby their carbonization is automatically accelerated. The smoldering bark, being usually decayed and dry, produces gases which come in direct contact with the specimens and are partially absorbed within their surfaces. Wood chips apparently possess the same qualities as bark, for the two types of fuel are used for this particular purpose interchangeably.

The following interesting statement appears in Myer (1928: 522):

> Mr. James Mooney, of the Bureau of American Ethnology, described to the author the following method which he had seen the Catawbas use in making their finest black ware: After the vessel or other object has received its final shape, and before it is baked, it is given a high polish by much rubbing with certain very hard and smooth stones or mussel shells with edges properly shaped by grinding. Over these unbaked, highly polished objects selected fragments

[55] Bushnell (1909: 12) records a somewhat analogous example in the treatment of pipes among the Choctaw: "When thoroughly burned [*i. e.*, fired] it [the pipe] turns rather dark in color, whereupon it is removed from the fire and immediately immersed in a bowl of grease, which is absorbed by the clay and carbonized by the intense heat."

of oak bark are piled, and the heap is then carefully and closely covered with a large inverted unbaked pottery vessel. . . . Over this unbaked pot a large amount of oak bark is piled and then set on fire. This produces considerable heat and bakes the large inverted vessel. The penetrating heat finally sets fire to the oak bark fragments underneath it, which, being shut off from a full supply of air, burn after the manner of charcoal and produce a strong, penetrating black, which reaches to a great depth into the ware, thus producing the beautiful color. The glossiness arises from polishing. The modern Cherokee produce a black which is much inferior to the above by burning ground corncobs in a small excavation in the soil, over which the vessel to be blackened is inverted. They also produce an inferior black by burning corncob meal within the vessel, which, in this case, is covered to prevent too rapid burning of the meal and the escape of the smoke.

The peculiar manner of firing presumably observed among the Catawba by Mooney should be dated to the last quarter of the nineteenth century. The method is neither practiced nor remembered by the contemporary Catawba potters.

One of the most recent innovations in firing is the use of the kitchen stove and a tin wash tub. Vessels to be fired are first placed within such a tin tub, reposing upon its bottom, some six to twelve pieces at a time. The tub is then set upon the stove, in which a moderate fire has been started in the meantime. More fuel is gradually added and the heat is increased, reaching 500° to 600° F. (260° to 315.56° C.). During this preliminary heating, the hearth to be used for proper firing is prepared, and the preheated specimens are transferred to it when they can be handled comfortably with the aid of rags. They are exposed to the heat of the hearth at a distance of about 0.8 m. from the edge of the ash bed. Only a few potters resort to preheating the vessels upon the stove.

Just what may be the reaction of the organic fat added in polishing, or in the process of firing, is not clear. Yet it seems reasonable to expect that its burning increases carbonization. When specimens so greased are completely smothered in fuel, the combination of readily burning organic matter (the fat) and the reducing atmosphere aid in carbonization and smudging. In open firing, a similar, though less generally constant, carbonization results from smoke contact or from smudging by embers; yet oxidation is naturally promoted by the free access of air. Whatever the causes, the blackish coloration is restricted, so far as its depth penetration is concerned, to the very surface of the wall (either in or out), or more properly only to the polished veneer. However, in the examples of roughly surfaced, corncob-rubbed, unsmoothed, and unpolished vessels which I saw fired at an open fire, the same type of blackening occurred, apparently from smoke contact; its penetration into the wall was again quite superficial. Inquiries revealed that the natives were not cognizant of the causes, although they fully realized the means of achieving the desired effect.[56] They were inclined to attribute the blackening to smoke alone, especially as they could point to an occasional puff carried by a leaping tongue of flame on to a vessel where a dark spot developed soon afterwards. They maintained that differences in the natural qualities of the clays, and also mere chance, were often responsible for the mottling.

It was not possible to obtain a satisfactory answer as to the criteria determining the completion of firing. With the pipes, which require just about an hour, actual testing is possible, for the specimens are small and can be handled with ease. In the case of large pieces, the changing color seems to be the guiding factor, although the resonance test is perhaps equally important.[57] The potter taps the vessel's rim with either a firm stick or a piece of stiff wire; at times she employs merely her hand, flicking a finger and striking the rim with the finger nail.

The initial step in firing, that is, the slow heating at a safe distance from extreme heat, brings about gradual drying. The chemical composition of the constituent clays then undergoes a reaction which is evident to the operator through the change of color. Under oxidizing conditions, the ferric elements produce a predominance of reddish or buff pigmentation; a reducing atmosphere, on the other hand, is largely responsible for the gray and black tones.

Finding the color and sonorous quality to her fancy, the potter removes the pieces from the hearth, or extinguishes the fire by disturbing the bed and heaping the ashes in the rear of the fireplace. As in all of the other steps of manufacture, experience and previous failure, unless a

[56] *Cf.* Speck (1934: 70).

[57] Speck (1934: 47, note 1) states: "*atuski*, 'pot, clock,' denoting the hour, is interesting in Catawba semantics. A clay pot rings clear like the stroke of an old clock when tapped on the rim; two o'clock becomes 'two pots, or rings,' and so on. Even the dove, *itusi* . . . earns a sobriquet from its call which resembles a stroke of the clock or a pot."

new experiment is pursued, seem to be the chief guidance of the potter. The specimens which I had an opportunity to watch in the firing process required from one to three hours of exposure to heat. I was told, however, that large pieces may need as much as twelve hours at the fire. As late as 1929 a tradition survived that none but the potter could witness the firing process, but it was not enforced. There are no records of this restriction published prior to that date. The geographically nearest similarity was noted by Bushnell (1909: 12) among the Choctaw.

The following temperatures, determined with the aid of an optical pyrometer, represent characteristic averages:

Open fire, out of doors	1670° F.—910° C.
Domestic fireplace	1760° F.—960° C.
Smothered fire in receptacle	1400° F.—760 °C.

A pronounced variety characterizes the color of the Catawba ware. Pieces fired under reducing conditions tend to be predominantly gray and black. The open fire, on the other hand, promotes reds and oranges. However, the majority of the ware is highly mottled; the colorful effect, as has been said, is considered an aesthetic asset; the mottling and the lustre have definite decorative values.

Attempts to express the numerous tones and hues on a single vessel in terms of a standard color scale are apt to meet with serious difficulties. The experience of Mrs. Fewkes and myself discourages me from placing on record the matching of some twenty Catawba pieces with Ridgway's plates (1912).[58] In one instance four basic tones were discernible (two in the grays, and two in the reds), while the mottled areas totaled forty-eight individual values. It seems adequate—and perhaps safe—to state that the basic colors of the Catawba ware range through the grays, oranges, and reds. Perhaps the most characteristic examples of the grays, as expressed in terms of Ridgway's nomenclature, include pale smoke gray, deep gull gray, mouse gray, and black. The reds and oranges, similarly expressed, comprise: avellaneous, cinnamon, English red, drab, grenadine, sepia, vinaceous, woodbrown. In cross-section, neutral gray and vinaceous buff are the most common colors.

[58] This work was done by my wife first; then a pottery preparator and I each made separate matchings. The three sets tallied about as much as might be expected from the limits of Ridgway's dependability as against a colorimeter.

The hardness of the Catawba ware ranges from 2.5 to 4, Moh's scale; the majority of the samples taken by me were under 3.5, and the average was 3. These figures apply to both surfaces and disregard the core as revealed by cross-section.

POST-FIRING TREATMENT

Once fired, the Catawba pottery is virtually finished. Some of the older potters either remember or still practice the smearing of hot vessels with a piece of fat (usually bacon rind or ham skin).[59] Such treatment produces a fine lustre which is particularly noticeable on dark tones; as a rule this process is restricted to the outside surface. Plain wiping with a dry cloth, especially of carbonized pieces, is often performed, although it is not always necessary. The cooled pieces, I noted, usually do not soil one's hands.

At times vessels are decorated, during the stage under discussion, with red sealing wax. This is done by applying a stick of wax over the outer surface of a moderately hot vessel and executing the desired, usually floral or simple geometrical, design. While I have not observed this apparently irregular practice, I have seen and examined about a dozen vessels displaying the sealing wax crustation. It is of interest, in this relation, to cite Gregorie (1925: 21):

> Mr. Phillip E. Porcher, formerly of St. Stephen's Parish, who lived to be more than ninety years old and died in Christ Church Parish in 1917, told me that he remembered frequently seeing the Catawba Indians in the days when they travelled down from the up-country to Charleston, making clay ware for the negroes along the way. They would camp until a section was supplied, then move on, till finally Charleston was reached. He said their ware was decorated with colored sealing wax and was in great demand, for it was before the days of cheap tin and enamel ware. This may account for the smooth, fresh fragments I have found on what are evidently old sites of negro quarters.

It seems safe to presume that the use of sealing wax for pottery decoration among the Catawba represents a definitely post-Columbian, and very likely quite modern, acquisition.

Whatever the post-firing care may entail, the chief purpose of handling the specimens immediately after their cooling arises out of a desire to examine each for possible defects. On the whole, casualties in firing are insignificant.

[59] Dr. Swanton (personal communication) has recorded the use of bear grease upon vessels among the Natchez.

There is no specially provided place for the storage of the final product in the fairly limited dwellings of the Catawba. The main concern, it seems, is to dispose of the output by selling as soon as possible. Merchants from Rock Hill formerly visited the reservation quite regularly to close the bargains, or the natives carried their ware to town and sold to dealers. The producer realized from five cents up to a dollar [60] and occasionally even more per piece, while the middleman cleared a handsome profit. In recent years, sales to shops at Cherokee, N. C., have reached a considerable volume, but have also forced prices at Catawba to a low level. At the same time the market at Rock Hill has just about ceased to exist.

TIME DURATION OF THE CATAWBA POTTER'S TASKS

The chronology of procedure, as here presented, is fairly constant. Such deviations as sometimes do occur are neither radical nor serious. The actual time element is a variable factor with the individual potter and the diversified steps incidental to the several processes. Clocking of the procedure is not always practicable.

The genesis of a pottery piece coincides with the departure in search of clays. The Catawba do not have to go far from home for pan clay; however, pipe clay is most readily obtainable in the river bottom and requires a ride of some 6 miles. The digging in a clay pit may entail several hours of labor; this includes the initial removal of undesirable matter. The mixing of raw clays takes perhaps an hour, and the shaping of the kneaded paste into lumps is a matter of mere minutes.

The preparation for the building process likewise consumes a short interval of time. With the modeling method, a vessel of a globular shape, with a maximum height and width of 0.1 m. and an average wall thickness of 0.005 m. can, I have witnessed, be shaped within 6 to 8 minutes. Since there is no scraping in such an instance, polishing and drying follow the building. In the stipulated example the specimen remained in drying, under an average temperature of 50° F. (10° C.), for approximately 17 hours; its polishing consumed about 8 minutes, and 2 hours later the potter pronounced the piece to be adequately dried for firing. The wiping, in which a dry piece of fine-textured cotton cloth was utilized, required about 2 minutes, and was accomplished during the preparation of a hearth for firing. A satisfactory bed of ashes was established in 15 minutes by burning sundry twigs, branches, and corncobs. The firing process of the specimen was completed within $3\frac{1}{2}$ hours. The particular example here described may be considered fairly representative of the several time durations experienced in the case of modeled pieces.

The ring and the circuit variants of building are, of course, more prolonged. The base disk is shaped in a very short time, the average being about 2 minutes, and the cutting of the strips of paste, out of which the fillets are fashioned, is done in 5 to 10 minutes. I had an opportunity to observe a ring-building example which may be tabulated as follows:

	Individual time (in minutes)	Compound time (in minutes)
1. Modeling of the disk (bottom)	3	3
2. Flattening of a lump of paste into an irregular slab about 0.01 m. thick, and cutting this into rectangles (repeated four times in order to gain adequate supply)	2	8
3. Cutting out 14 strips, and placing them separately upon an auxiliary board (12 to be shaped into rings, the remaining 2 eventually to become handles); average dimensions of strips: length 0.28 m.; width 0.015 m.; thickness 0.01 m.	1	14
4. Rolling the strips into fillets, and adjustment of their length and girth to uniformity; dimensions: length (average) 0.295 m. (increase over strip due to rolling), diameter (average) 0.015 m.	1	12
5. Shaping and closing each individual ring	1	12
6. Applying, adjusting, and bonding the initial ring upon the base disk	3	3
7. Application, adjustment, and bonding of each of the subsequent 10 rings	1	10
8. Placement, adjustment, and bonding of the terminal (twelfth) ring to form the rim and lip	2	2
9. Smoothing, by hand, the entire body of the cylinder inside and outside		4
10. Shaping of the form, producing a jar globular in the lower portion of the body, with a gently profilated shoulder, and a cylindrical neck		14
11. Scraping; inside and outside (with a cane knife)		9
12. Smoothing the rough surface with moistened hands and with a rag soaked in water		2
13. The vessel was next set aside for drying on the mantelpiece; the average temperature of the room was about 60° F. (15.56° C.); the vessel remained in		

[60] Speck (1934: 71) records 25 cents.

drying for 4½ hours. Thereupon, the resumed operations entailed:		
14. Pebble polishing, over the entire outward wall, bottom, and approximately three-quarters of the height of the neck on the inside		13
15. Appending of 2 diametrically opposed, flattened, sharp-edged rod handles, akimbo in appearance, each attached on the neck and shoulder respectively, utilizing 2 of the 14 strips of paste (see step 3), but modeling the shapes after attachment	7	14
16. Final adjustment of rim by flattening its lip		1
17. Restoration of pebble polish in areas affected by the attachment of the handles		2
18. Polish with a piece of bacon fat		2

The specimen was now in its completed plastic stage.

The handles, made of the two fillets which during the drying interim were placed under moistened paper, required about 4 hours to be freed, by normal evaporation, of the bulk of their admixed water. The partial drying incidental to their storage despite the proper protection, was dismissed by the potter as inconsequential.

Thus the total compound time up to this stage amounted to 2 hours and 5 minutes.

Preparation for the firing process, with all the necessary prerequisites and accommodations close at hand, was a matter of a very few minutes. The ash bed was formed in approximately half an hour, with fuel similar to that cited in the instance of the modeled piece.

The vessel under description, and other pieces previously made, were examined for defects and, being found satisfactory, were placed on the floor before the fireplace at a distance of about 1.5 m. There they remained exactly 40 minutes, being turned, in the meanwhile, about every 10 minutes. Then they were advanced within 0.7 m. of the ash bed, allowed to be examined by me in the process, and having been turned at intervals of approximately 5 minutes each, for fully 1 hour, were finally pushed practically into contact with the frequently fueled flames. There they were permitted to remain, with still additional turning, at an average frequency of 20 minutes (the rake being called upon for the purpose), for 2 hours. The computed time of the graduated firing, therefore, totalled 3 hours and 40 minutes. Upon the completion of the firing, the glowing embers were extinguished by raking and the ashes were heaped into the rear of the fireplace. Thereupon the vessels began to cool. At the end of half an hour they were hot; in another 15 minutes they were quite warm, and 10 to 15 minutes later they could be handled comfortably. Certainly the maximum time necessary for their complete cooling did not exceed 1 hour; the room temperature averaged about 68° F. (21.11° C.). The time consumed in the examination of each specimen then varied from 1 to 3 minutes, the particular vessel here described requiring little over 2 minutes to be pronounced entirely satisfactory. As the final touch, the potter rubbed the batch with the same piece of bacon fat previously used by her in culminating the pre-fired surface treatment; this took but 5 minutes of her time.

Mrs. Sally Gordon, the operator in both cases here timed, was a very skillful and speedy potter. Her dexterity, efficiency, and form tastes were, I think, distinctly above the average. The record of Mrs. Sally Gordon, proudly revealed by herself and corroborated by others, of having completed eight vessels by ring building, two by modeling, and adding a half dozen molded pipes, all in the span of a single working day, represents an outstanding achievement.

Two hours seem to represent a fair average for the construction and surface finish of a simple form; attachment of appendages, curvatures in shape, etc., may call for an additional mean of about 30 minutes. This estimate is equally applicable to the ring and the circuit modes of building.

The sectional building process approximately doubles the time necessary for the uninterrupted procedure. The largest, and most complicated vessel ever so produced within the memory of my informants, was said to have arrived at its final state of construction after some 6 hours of uninterrupted work.

A pipe blank can be modeled by hand within 5 minutes. The molding of a pipe in a double form is usually a matter of 10 to 12 minutes (fig. 8). In either case the surface treatment is partially achieved concurrently, and if done entirely separately, it requires no more than 8 minutes. The decoration is more tedious, the plotting of the motif taking as much as 20 minutes; yet the actual execution of the design can be accomplished within half that time. Drying takes one day or less, the carving of the tobacco bowl and the perforation of the stem for smoke passage some 20 minutes, and firing about 1 hour, while the final polish is often done in as little time as 1 minute.

TECHNOLOGICAL MISCELLANEA

This section is concerned only with certain *mechanical* aspects of technology. Such aspects pertain primarily to the constructional, surfacing, and finishing processes. On the other hand, those approaches which depend chiefly on ceramics, chemistry, optics, petrology, etc., may be collectively designated as analytical technology. Naturally, the domain of ceramic technology—unsatisfactory as the term itself may be—should bring about a co-ordination of the two subdivisions. Its subject matter is so involved that it defies the control of any individual, and yet a high degree of overlapping must inevitably be recognized. It is a fortunate circumstance that the American specialists in pottery technology, representing a variety of qualifications and interest, have already reached a division of tasks. This was demonstrated by the conference held in November, 1938, on which a brief note appeared in *American Antiquity* (**4** (4): 358–359, 1939).

With the Catawba, true coiling is totally unsubstantiated either in literary sources or in the recollection of those informants whom I interrogated. (When the potters themselves refer to "coiling," they actually have in mind the annular variants of construction.) The archaeological sherds collected and examined by me at the reservation revealed no evidence of coiling. It is interesting to note, in this connection, that Catawba potters settled amid the Cherokee, with whom coiling unquestionably formerly was really characteristic, appear not to have adopted this method. That such women may well have seen Cherokee potters use coiling must be admitted as a reasonable certainty.

The circuit and the ring variants, the Catawba insist, offer equally convenient means of wall-building. Both are commonly used, and are said to be equally old. With respect to the validity of this contention, it may be stressed that witnessing of the two variants convinces the observer of their equal expediency as to manipulation and speed. The construction of a blank cylinder, subject to subsequent form-shaping, a characteristic phenomenon at Catawba, is also believed by the natives to be an old practice; indeed, it appears to be preferred to tecto-shaping. It seems permissible to presume that certain curvatures might easily be attained—consciously or unconsciously—in the process of fillet-building, depending on the circuit variant or even perhaps on the rings. During my observations I did not notice any deliberate attempt to do so.

The cylinder itself, as has been stated, is a characteristic peculiarity in the Catawba building technique. Its several advantages are particularly favorable to the separate step of shaping the form. The dimensions of the finished form often radically differ from those of the blank. By way of a concrete illustration of this point, I may return to the example for which time equivalents in construction are tabulated in the preceding section. The cylinder in question, when completely erected, had a maximum height of 0.135 m. (Theoretically, the maximum of the superimposed but unpressed rings, twelve in number, added to the base disk 0.015 m. thick, and the adjustment of the rim, compute to a total of 0.15 m. The factual measurement of 0.135 m. reflects the alteration, *i. e.*, compression, effected by the bonding.) The body of the cylinder was fairly regular in curvatures; when its erection was completed, the mean mouth diameter measured 0.083 m. (reading on the inner side of the rim), whereas the mean maximum diameter of the bottom, taken on the outside, amounted to 0.105 m. The fired vessel has the following maximum measurements: height, 0.153 m.; diameter of body (outside limits), 0.104 m.; diameter of bottom, 0.065 m.; diameter of neck at its junction with the shoulder, 0.075 m.; mouth diameter (outwardly), 0.065 m.; thickness of wall (taken at the lip), 0.004 m. (fig. 15*a*).

This same vessel has a grayish buff color, and approximately one-half of its outside surface has mottled areas in varying shades of orange and gray. Their distribution begins on the lip, runs over the neck, handles, and belly, and also the bottom. The variegated effect is due to oxidizing atmosphere, to reducing gases, and to smoke carbonization. The upper inside margin of the neck is only faintly discolored in blackish gray in places adjacent to the lip mottling. The bottom shows an undiscolored spot, some gray and black tones, and a most instructive streak of a slightly darker shade of the basic color. The streak, darkened by the carbon of the smoke rolling up the wall, is also plainly visible on the belly through a similar, although smaller, zone, which indicates where the smoke stream started. While reducing conditions prevented mottling of the streak on the bottom, oxidation stimulated it on the exposed belly. The smoke was derived from several slow-burning clinkers expelled from the fire and settled near the vessel (which re-

mained in an upright position throughout the firing); the blackening was then due to reduction and acute carbonization. The streak on the belly, however, was partially subjected to additional penetration of smoke and was somewhat darkened thereby; a close examination reveals the extent of its original distribution. The darkest tone of gray appears on the lower portion of the belly, off the base. In watching the firing, I noticed that a well-glowing clinker became wedged there and promoted the process. In the areas of its closest contact with the vessel, the basic color was not changed. However, marginally, and only vertically up the wall, excessive blackening resulted, being quite sharply delimited by an undulation in one spot, and rather blurred elsewhere. The severe demarcation appeared to be incidental to two factors, namely the shape of the contacting portion of the clinker, and the concentration of reducing gases immediately above it. The smoke found freedom in rolling upward; the soot settled upon the surface of the vessel in a broad span running over the lip and also reaching the inner side of the neck. I suppose that the heat rising from the interior of the vessel prevented its deeper penetration inwardly down the neck. The same rolling column of smoke also left traces on the sides of the handles facing its course; however, certain deflection took place there as evidenced by the discoloration of the horizontal bars of the handles. The handles have their edges defined by unequal sharpness and angularity. It is interesting that in the one having a more or less rounded form, the smoke coloring continues quite gradually on to the planed underside. The angularity of the other, however, all but stopped this distribution. A similar phenomenon is visible on its laterally opposite edge, but is due to another batch of smoke. Elsewhere on the body other splotches, each the result of smoke contact, are present.

Despite the knowledge gained from observing the processes just described, certain elements operative in mottling remain unexplained. Outstanding among these are the common examples of concentric shading within a given spot; the formation of peculiarly distributed areas on the interior of a vessel; and the frequent occurrence of spattered, small blotches, apparently occasioned wholly by smoke rather than by contact with individual clinkers. Whether or not it is possible to rationalize these phenomena inductively remains to be determined. The point should be pursued by further inquiry; it is possible that high-speed photography and special filters might be helpful in such an endeavor.

Similarly, the use of blood or fat in the preparation of paste, and their full significance and value, are incompletely understood. As already stated, I did not see the practical application of such media during my studies.

CHEROKEE POTTERY-MAKING

It seems appropriate at this time to consider the present status of pottery-making among the Cherokee, whose industry has been under Catawba influence for a considerable period of time. If it is permissible to speak of acculturation in this individual instance alone, then the phenomenon involved illustrates an acute example of its practical, dynamic manifestation. Now that imported pipe and pan clays from Catawba have replaced the allegedly depleted local supply, the superficial resemblence between the two categories of pottery is very close indeed.

Mooney (1900: 164) spoke of Ewi Katalsta as "the last conservator of the potter's art among the East Cherokee." Her process was observed as late as 1908 by Harrington (1908*b*: 222 ff., pls. 2–10).[61] It was previously described by Holmes (1903: 56) who obtained his information from Mooney; and Holmes (1903: 53) also wrote of Catawba potters among the Cherokee, likewise after Mooney. One of these, named Susan, the wife of the Cherokee ex-chief Sampson Owl, was a constant maker of pottery, entirely in the Catawba style, for over forty years;[62] she died in 1934.[63] Her daughter, her relatives, and others who desired to learn the craft, were readily taught by her or copied her procedure.[64]

Ewi Katalsta[65] was her contemporary; it can be demonstrated, I think, that even she was not immune to influences of the Catawba potters living at Cherokee. That she could have adopted certain elements from Catawba immigrants at Cherokee is, of course, wholly within the realm of probability. Moreover, it is possible to point to

[61] Harrington stated that there was, in 1908, only one other potter at Eastern Cherokee, "an aged woman known as Jennie Arch, whose feeble hands had all but lost their skill." He apparently was not cognizant of Susan Owl; *vide infra*, and Speck (1939: 25–26).

[62] Personal information from Dr. Speck.

[63] The date is apparently only approximate; *cf.* Speck (1939: 26).

[64] Personal information from Dr. Speck.

[65] Ewi Katalsta was still a somewhat active potter as late as 1913; she died about 1926. (Information from Dr. Speck.)

concrete evidence that this was actually so. Mooney observed that "in building she [Ewi Katalsta] *sometimes*[66] used one long coil which was carried spirally from the bottom to the rim after the manner of the ancient Pueblos and the Louisiana potters."[67] There is no explanation as to which mode of construction she employed *otherwise*. Conjectures are unnecessary, for Harrington's account of Ewi Katalsta's building furnished conclusive proof of its analogy with the circuit variant of the Catawba: she applied the paste roll, and adjusted it, "pinching it fast the while *until the circuit was completed*. The *coil* proved too long, so she broke the superfluous piece off and blended the two ends together with care. . . . Thus the coiling proceeded until the required form and height were reached."[68] Note also the omission of temper in the instance witnessed by Harrington (1908*b*: 224), which is yet another Catawba rather than Cherokee characteristic.

In 1929 Dr. Paul Kirchhoff collected an old vessel in the Cherokee country,[69] which seems inferentially attributable to the work of Ewi Katalsta.[70] The specimen (fig. 30) embodies a debased Cherokee form and decoration (grooved, pointed garlands, each formed by four parallel lines, with a fringelike effect, vertically placed, and a similarly executed cross-hatched design on the neck, which is slightly raised above the shoulder) and typically Catawba polishing.[71] The shape of its body approximates presumed ancient Cherokee forms,[72] and the execution of the decoration resembles similarly ornamented sherds thought to be of Cherokee provenience.[73]

FIG. 30. Cherokee vessel, probably the work of Ewi Katalsta. Collected by Dr. Paul Kirchhoff.

I am not acquainted with analogous comparisons in so far as the decorative motif and the rim treatment of the vessel collected by Dr. Kirchhoff are concerned. At any rate, the specimen, whether the product of Katalsta or not, illustrates perhaps one of the last flickers of the now vanished Cherokee work based largely on their own traditional expression. Although it is not possible to judge the building technique accurately from surface appearance, a certain amount of concentric alignment, not all due to polishing, suggests an annular method. The present concern with Katalsta may well be terminated by recording Dr. Speck's significant statement that "by 1913 Ewi was no longer a potter purely in the Cherokee style, but markedly influenced by Catawba methods."[74]

At the present time Catawba-type and Catawba-inspired ware is being produced by the Cherokee as a permanent source of revenue. The tourist trade, served by "craft shops" on the reservation and elsewhere, provides the chief market. In season the demand is quite lively, and the local potters are unable to produce an adequate supply. The shortage is made up by the Catawba, who either bring their products to Cherokee in cars, or sell them at their reservation to visiting shopkeepers and agents from western North Carolina and elsewhere. Thus Catawba

[66] Italicized by the present writer.

[67] Holmes (1903: 53). (The extraterritorial comparison may have been drawn by either Mooney or Holmes; that, however, is irrelevant to the point under consideration, which applies to the Cherokee potter.)

[68] Harrington (1908*b*: 225); italics by the present writer. *Cf.* Holmes (1903: 53–54) for identical procedure by two Catawba women at Cherokee in *1890*; Holmes's index (p. 207), not text, classifies this as *coiling*.

[69] Specimen originally deposited in the Department of Anthropology, University of Pennsylvania, where I have examined it.

[70] Dr. Kirchhoff accompanied Dr. Speck and myself to the Catawba reservation prior to his trip to the Cherokee. (The vessel in question impressed him as a valuable specimen, and he investigated, without positive success, its history; personal information from Dr. Kirchhoff.)

[71] I am obliged to Dr. Speck for permission to refer here to this interesting specimen.

[72] *Cf.* Harrington (1922: pls. XLIX*d* and LII; also pls. XLIX*g* and L, both from Tennessee).

[73] Harrington, 1908*b*: pl. LVII*a* (particularly the lower three scrolls of the left portion of the design), *d*, and *f*.

[74] Personal information from Dr. Speck.

pottery is often sold as a Cherokee product, or farther away from the Smoky Mountains National Park, simply as Indian pottery. To one acquainted with the characteristics of the Catawba ware, the identity is immediately obvious no matter what the label. It seems appropriate to note that some Cherokee potters have been approached by commercial buyers from Minnesota and Arizona. However, there has been no favorable reaction to such offers, perhaps chiefly because the local market is so good. A proficient potter at Cherokee, according to local information, may realize as much as twelve hundred dollars on sales of her ware within a single year. Such an income, almost entirely clear profit, provides a revenue which, in view of local standards and economy, is quite considerable if not extraordinary. (At Catawba, an entire family of potters with four expert women at work and one or two men to dig clays, does not make six hundred dollars in a year.)

As late as 1934, Dr. Speck collected three vessels made by Mrs. Maude Welch, a native Cherokee and an accomplished potter, of which two are of the Catawba category and one in quasi-old-Cherokee style, paddle-stamped.[75] The paddle-stamped specimen is quite inferior in fabric and general execution as compared with the two Catawba-inspired pieces made by the same potter. Truly, the latter two pieces rival the average ware now manufactured at Catawba. In 1941 Mrs. Welch informed me that she used local clay in making the stamped pot, and Catawba clays (pan and pipe) in the other two pieces; and that she constructed all three vessels by the circuit method.

FIG. 31. Mrs. Maude Welch (Cherokee) working on a pot on which Indian heads with braided hair are used as lugs. A pair of toothless combs, a spoon, and a knife are among the tools upon her working board.

The contemporary Cherokee no longer manufacture any of their own original type of pottery. Similarly, vessels distinctly characteristic of the Catawba quality and technique, and yet combining such native Cherokee elements as coiling and the use of the paddle, are also not made any more.

Mrs. Lillie Bryson (*cf.* note 2) continues to make typical Catawba-style pottery which is sold as Cherokee product. She, like Mrs. Welch, uses Catawba clays; moreover, her former experiences at Catawba still tend to dominate her work. She is, therefore, less inclined to deviate from old tradition, although not always successful in resisting certain new temptations.

With the express permission of Dr. Speck, I should like to mention a particularly interesting and instructive pot manufactured by Mrs. Bryson late in 1935. This is a vessel of a globular body with two independent necks. The necks are surmounted by a ribbon handle forming a high loop and attached to the mutually opposed margins of each rim. In 1935 Dr. Speck was informed by Mrs. Bryson that the specimen was originally conceived by her to serve the purpose of a wedding gift; her idea, it seems, was motivated by a desire to symbolize the union of the intended recipients. Mrs. Bryson reiterated the substance of these statements in answer to my inquiries in 1941. However, my further pursuance of the subject—since the "wedding pitcher" is now quite common both at Cherokee and at Catawba—revealed a significant factor. In 1935, at an Indian fair held at Atlanta, Georgia, several Pueblo potters are said to have been present, among them the renowned Maria Martinez of San Ildefonso—at least that was the name my informant was able to recall with some help on my part. In any case, the Southwestern

[75] All three specimens originally in the Department of Anthropology, University of Pennsylvania, where I have examined them with Dr. Speck's permission to mention them here.

artisans demonstrated their native pottery-making, and the vessel with two necks, a normal Pueblo form, was among the finished products. Mrs. Youngbird, a Cherokee potter, attended the fair, and it appears quite certain now that she is to be credited with the introduction of this Pueblo form at her reservation. Mrs. Bryson's claim to independent invention of the pitcher with two necks has a counterpart at Catawba. And yet it was possible, by following chronological evidence and trade contacts, to reconstruct the actual course of events. Catawba potters, visiting Cherokee to sell their ware, noticed the new form *after* 1935 and promptly adopted it as an expedient economic advantage.[76]

Mrs. Bryson, now living at Ela, 6 miles away from the Cherokee reservation, is a very active potter. Her products are made from imported Catawba clays, and upon the principles of Catawba techniques from start to finish. They are easily recognizable in the several "craft shops," and can be fully identified by the maker's name and affinity appearing on the bottom. Mrs. Bryson has added the pitcher with two necks and the frog pot to her forms since she took residence at Cherokee in 1930.

Dr. Speck secured yet another interesting vessel during his 1935–1936 work at Cherokee. This is a bowl with a frog figure in positive relief enveloping the upper portion of its body,[77] recently made by Mrs. Maude Welch. The occurrence of the applied frog motive was *presumed* amid ancient Cherokee (?) pottery by Harrington (1922: 283).[78] According to her own statement, Mrs. Welch arrived at her idea through observation of archaeological specimens in the B. S. Colburn collection at Biltmore Forest, North Carolina.

[76] The vessel with two necks made by Mrs. Bryson in 1935 has, in the meantime, been published; *cf.* Pennypacker (1937: 147–148, fig. 1). However, Pennypacker gave no account of its then alleged history; his statement that "archaeologically these pots have a very wide distribution which extends from Canada to Florida along the Atlantic Coast" is certainly contrary to truth.

[77] The vessel is somewhat analogous with the specimen shown in Harrington (1922: pl. LIX), but the modern piece has the frog more pronouncedly in relief. I am indebted to Dr. Speck for permission to refer here to this vessel.

[78] Such presumed Cherokee provenience of the frog pot amid archaeological remains in the Southeast cannot claim any recognition, since there is no positive evidence to support it. The frog is not represented among the pottery remains recently found in historic Cherokee sites in Tennessee, (personal information from Miss Madeline Kneberg).

Mrs. Welch is doubtless the most skillful native Cherokee potter at the present time. She learned her craft at Catawba, where she still visits occasionally, and is, on the whole, rather conservative. The frog pot, and a conoid jar with applied snake about the body or with two Indian heads provided with braided hair (fig. 31), are typical and ever readily recognizable on any store counter. (They are usually identified by her name and residence inscribed on the bottom.) The local trade prefers her products, and they bring higher prices. Mrs. Welch makes relatively few pitchers with two necks, which she views as outside intrusions not in line with domestic tradition. She obtains her clays from the Catawba reservation, uses two parts of pipe clay and one part of pan clay in making vessels, and employs the circuit variant of annular fillet construction. She resorts to modeling only in producing small pieces. Mrs. Welch relies rather heavily on the use of a moist rag for smoothing, reducing wall thickness, and even some shaping, as well as polishing. Her tools depend more on modern kitchen pieces than on native means; mussel shell, gourd, cane knife, bone polisher are not included in her tool assemblage. On the other hand, the polishing pebble is of prime importance, also a cherished possession, and usually quite old. Mrs. Welch employs modeling in making pipes, the Indian heads, and the snakes and frogs for appliqué; the blanks are quite crude in either instance, but carving with a metal knife-blade and smoothing with a wet rag eventually lead to the execution of delicate details in many cases. The applied features are in all cases solid, irrespective of dimensions; special care is taken in their drying process, which is invariably prolonged, and in the firing, which is done very slowly and under minutely controlled increase of heat. Mrs. Welch fires her products at an outside fire and as a rule preheats them in a washtub on the kitchen stove.

Aside from Mrs. Welch, there are some five or six native Cherokee women on the reservation who make pottery more or less occasionally. Their standards and skill are not comparable to those of Mrs. Welch and Mrs. Bryson. Moreover, it appears that these unstable potters are responsible for some of the most grotesque products at Cherokee, such as the figure of a squaw with a papoose on the back, imitations of modern industrial pottery forms or their various features, etc.

The "craft shops" at Cherokee offer the gulli-

ble tourist numerous gift "attractions" and appear to be doing a thriving and profitable business. It is quite natural, then, that the pressure of commercial opportunities exerts its influence upon the local sources of supply. The potters are exposed to a veritable motley of temptations to which they succumb, either of their own will, or upon an impressive sales talk, rather easily. The result is often quite ghastly from the ethnological standpoint, although it may well delight the average tourist. The shopkeepers are interested in the profit, and the natives appear to be oblivious to the potential ultimate consequences of abandoning old tradition. As a result, truly native pottery-making no longer exists at Cherokee.

The agency school had an opportunity to salvage and resurrect the craft during the last decade, but the attempt actually made was a failure. Instead of stressing native methods, the potter's wheel and a kerosene kiln were introduced. Photographs of archaeological vessels from the Colburn collection were used as models despite the circumstance that they did not include a single specimen of true Cherokee affinity. Apparently the lack of a competent instructor, acquainted with the history and technology of native Cherokee pottery-making, was a particularly serious handicap. In any case, the abortive endeavor to attract and stimulate local interest made no impression on the active potters, although some of the artisans participated in the initial work. The school authorities are anxious to secure a properly qualified instructor, and it is to be hoped that the significance of retaining native procedures will not be overlooked. The clay shortage is to be met by importation of suitable supplies to be sold, under government supervision, to local potters at cost. It should certainly be expedient to obtain an ample clay supply at Catawba. And it cannot be urged too strongly that the wheel and kiln be relegated to oblivion in the contemplated program. The history and tradition of Cherokee pottery lend themselves well indeed to rational vocational training.

As has been stated, pottery-making at Cherokee has lost the truly native procedure in which aboriginal methods predominated. And yet some specimens may even now be found on the reservation which, although quite inferior in execution and fabric to the old style, retain certain elements of native technique. Such specimens are recognizable by the admixture of crushed pebbles for temper, often by the use of the coiling method, and by the carved wooden-paddle stamping or textile-wrapped paddle impressions;[79] the blackening of the interior, achieved by the firing of readily burning fuel,[80] such as bran,[81] corncobs, or tree bark; the seemingly lesser degree of care given to clay selection, which stands out in particular contrast to the Catawba practices;[82] and by the surface color, which is almost entirely buff (on dull ground) and only rarely blackish (on lustrous ground). Holmes (1903: 56) viewed "the application of a black glossy color by smother-firing" as one of the chief differences between Cherokee and Catawba wares. According to Harrington (1922: 196), the native Cherokee processes "probably are practically the same as in prehistoric times." This assumption is quite vague in view of the author's own concept (1922: 174–175) of an early group and a later group of aboriginal Cherokee pottery. Harrington does not give the manufacturing technique for either group. Holmes (1903: 161) inferred, although not conclusively, the "coiling" method.[83] Mooney was informed by a Catawba woman at Cherokee that "the manner of baking by which a rich black color was given the ware was . . . acquired from the Cherokee" (Holmes, 1903: 53).[84] As far as possible interchange between Cherokee and Catawba potters in earlier ("prehistoric") times is concerned, there is no sufficient ground for deductions.

The pottery now being made at Cherokee comprises pieces produced by transplanted Catawba methods, either locally taught by Catawba immi-

[79] *I. e.*, purposely applied for decorative purposes, not incidental to the shaping manipulation.

[80] Probably also practiced by the ancient Catawba.

[81] *Cf.* Harrington (1908*b*: 226) and Holmes (1903: 56), where "wheat or corn bran" is specified. This method was occasionally used by Mrs. Welch even in 1941.

[82] This peculiarity is perhaps explainable as being due to the shortage of suitable clay. According to Mrs. Welch, the best deposits, situated on Soco Creek (*cf.* Harrington (1908*b*: 224)), became unavailable about twenty years ago, when some buildings were erected upon the clay beds.

[83] *Cf.* also Holmes (1903: 50–51) for the assumption that in the Eastern United States "the strips of clay were wide, irregular, and rude, and . . . rarely showing traces of their employment." Yet, immediately thereupon, Holmes stated: "Specimens from many sections of the Eastern United States fracture along the strip junction, thus revealing the fillet, and the manner of their manufacture." It may be recalled, in this connection, that Catawba women settlers among the Cherokee appear never to have adopted the true coiling technique.

[84] Fuel-smothered firing, reducing conditions, and controlled mottling are all commonly employed at Catawba to attain black or dark gray tones; the methods involved appear to be survivals of old practices.

grants or acquired by Cherokee women at the Catawba reservation. Except for certain modifications resulting chiefly from lesser general skillfulness, and some differences in form, this pottery resembles that of the Catawba reservation. To one familiar with the latter, however, the inferiority of the Cherokee product, qualitatively speaking, is quite obvious. The surfacing and the color scheme of the Cherokee-made pottery do not equal the aesthetic effects attained by Catawba women at their home. Since Catawba clays are generally used now, the basic similarities in outside appearance should be rather close. And yet there are clear-cut differences even between the products of Mrs. Welch, the best native potter, and those of Mrs. Bryson, who originally came from Catawba.[85] It appears that at Cherokee, within the last decade in any case, the potters were confronted with a lively and gradually increasing demand for their vessels. To meet the situation, and thus properly to capitalize on the opportunities, production had to be accelerated. Consequently, some of the details of technique suffered and the general care became rather superficial. Although promising local talent clearly manifested itself from time to time—as exemplified by Mrs. Welch and others—its potentials evidently did not stimulate any craft consciousness among the artisans. Therein lies the fundamental difference between contemporary Cherokee and Catawba potters. The Catawba women, and indeed the "nation" at large, are proud of their craft; they derive a definite sense of satisfaction from practicing it, and its economic value, although an asset, is certainly modest. The Cherokee had lost their own native methods over two decades back when Ewi Katalsta ceased to be active. They subsequently resumed pottery-making through Catawba tutelage, which became more pronounced than ever before. Ultimately, Catawba technical dominance became supreme at Cherokee, and continues to be to this day. However, the Cherokee never quite grasped the finesse which characterizes pottery-making at Catawba. Perhaps there has not been adequate time yet for a thorough assimilation. Whatever the rationalization, the circumstance cannot be overemphasized that the potter's craft at Cherokee, as now practiced, is nurtured primarily by the thriving tourist trade.

Forty years ago Holmes wrote (1903: 142): "It is not yet possible to make a satisfactory analysis of the pottery of the Carolinas." Referring to the pre-Columbian ethnic complexity of the area, he also stressed that the field has been "little studied," a point which remains true today. In extant literary sources information on Catawba and Cherokee pottery technique is sadly deficient throughout the period antedating the observations of Palmer and Mooney. For the Cherokee, Timberlake (1765: 62) merely mentioned their "two sorts of clay, red and white, with both [of] which they make excellent vessels, some of which will stand the greatest of heat." Butel-Dumont (1753, **1**: 154; **2**: 271 ff.) and Du Pratz (1758, **1**: 124; **2**: 178 ff.) recorded certain eighteenth century practices in Louisiana. On the other hand Lederer (1672), the first European to publish about the Catawba and the Cherokee, did not mention their pottery at all. A century later Smyth (1784: 193) referred to Catawba pottery as "an ill-formed kind of half-baked earthen ware."[86] This statement, obviously, is valueless for the present purpose.

Commenting on modern Cherokee potters, Holmes (1903: 143) described them as skillful and stated:

> Their ware has several points of analogy with the ancient stamped pottery of the South Appalachian province. Their ware retains more of the characteristic elements of form than does that of the Catawbas and the stamps they use in decoration are identical in many respects with those found used in the entire region extending from southern Florida to Virginia.

(It is to be remembered that Holmes did not observe either Catawba or Cherokee potters at work. His judgment was based on notes taken by Palmer and Mooney, and on vessels collected by them. His lack of field observations was a serious, although not consciously recognized, handicap.) And referring to former Iroquois manufacture, Holmes wrote (1903: 161): "Evidences of the building process are obscure, but there is no reason to suppose that other than the *usual methods* were employed."[87] Cushing (1894:

[85] The products of Mrs. Welch and Mrs. Bryson can be readily recognized in the motley groups of pottery for sale at various shops. The respective names of the two potters appear on the bottoms, so that precise identification is possible.

[86] Bushnell (1906: 673) quoted this passage from Smyth without any comment.

[87] Italics those of the present writer. Obviously, Holmes's negative comment has no specific value. The obscurity in question is not to be minimized in so far as complete, undamaged vessels are concerned, although even in such cases certain distinguishing criteria of technique are often discernible.

227) considered Cherokee pottery-making more difficult and tedious than that of the "shore people," but did not substantiate his contention.

PAMUNKEY POTTERY-MAKING

Formerly certain constituent tribes of the Powhatan Confederacy (in the sense of post-British contacts) had a "smooth" ware which resembled, as Speck (1928: 411 ff.) has shown, that of the Catawba. The Virginia category of this pottery, differing from that of the Carolinas chiefly in its powdered mussel-shell temper, is perhaps to be assigned to the Chickahominy, the Mattaponi, and the Pamunkey (Speck, 1928: 339 ff.).

> The smooth ware which finally usurped the style and technique at Pamunkey was known to the natives of much of the east. Sherds of the same texture and surface are found in the Cherokee region, among the Catawba, and all over the tide-water Algonkian habitat from the North Carolina-Virginia boundary to the head of Chesapeake Bay. We have specimens to illustrate this from the Chickahominy through the country to the Nanticoke area of Delaware. (Speck, 1928: 411–412.)[88]

The nineteenth century methods of the Pamunkey potters, who appear to have abandoned their manufacture for local needs with the advent of the railroad in years antedating the Civil War (Speck, 1925: 409), were observed and described by Pollard (1894).[89] Drawing on this record, Holmes (1903: 153) added certain comparisons, namely that "this pottery corresponds somewhat closely in general appearance with that of the Cherokees and the Catawbas."[90] Harrington (1908*a*: 406) referred to modern Pamunkey pottery as "crude attempts for the curio hunter."[91] However, twenty years later Speck (1928: 409 ff.) presented a convincing account of the traditionally remembered potter's craft at Pamunkey.[92] This conclusively proved Catawba derivations, which Speck (1928: 414 ff.) traced historically. Especially significant are his details regarding the initial preparation of the basal disk, the building of the wall, which was done by "adding thin layers of clay paste," direct shaping by modeling, and also his statement that "the coiling was not followed in recent times" (Speck, 1928: 410–411). All this represents a new contribution. Speck's description (1928: 411) of the method of scraping, pebble polishing, and fuel-smothered firing corroborates the previous observations of Pollard (1894: 18). The latter author, however, spoke of "kiln baking," without explaining the contraption.[93] It is possible that the Pamunkey arranged their fuel in a manner which may have impressed Pollard as a kiln. Speck's specific description (1928: 411) induces an acceptance of such a possibility: "Next comes the burning of the pots in the open fire hearth. The Pamunkey cover the jars with corn-stalks and pieces of dry pine to give them a light-gray color. The stalks and bark are piled over them to cover them in burning."

Speck's comparisons of modern Pamunkey and Catawba methods should be applicable, in view of his genealogical studies, as far back as the first quarter of the nineteenth century; this deduction led him (1928: 418) to the following conclusion: "Would it not seem plausible, then, to ascribe an early manufacture of the smoothware to both surviving groups?"

At Pamunkey, according to Speck (1928: 402–404, figs. 101, 102):

> Its sporadic occurrence, its localized abundance, and some historical circumstances, as well as the ethnological conditions among the present Indians of the region, point clearly to the conclusion that the ware of this type came into being after the natives had changed their economic habits resulting from contacts with the English. . . . The ware is characterized by being very smooth, hard and fine-grained, the clay free entirely from sand and grit, yet full of powdered mussel-shell. Its color is light brown or uniform drab or gray. No incised or depressed decorations are found in the body. A few rims only show any attempts at embellishment, which then consists of fine impressions or dents, sometimes of finger marks. Next is the most important thing: numerous angular bottoms, parts of curved handles or lugs, legs and knobbed lids, together with evidence of flat bottoms and the exclusive lipped rim style, are indications of a modification in form, bringing them into correspondence with the common European forms.

[88] *Cf.* also Holmes (1889: 249–250).

[89] *Cf.* also Mason (1887).

[90] For illustrations of Pamunkey specimens (collected by Dalyrimple about 1878 (Holmes, 1903: 152)), *cf.* Holmes (1903: pl. CXXXVI), in which analogies with Catawba forms may be seen in the two plain bowls (front row, left and center) and the three(?)-footed mug (rear row, center, shown with a separate bowl reposing within its mouth).

[91] Here, however, it is also stated that the Pamunkey "still make a few earthen pipes, some of which are of old form and all of which, I understand, are made by old time method to a great degree."

[92] Vessels and pipes illustrated in his figures 108 ff.

[93] Holmes (1903: 130) rejected the sporadic reports of kilns in the Southeast.

Here then is the secret, and, comparing this material with the historic Pamunkey ware, we are forced to conclude that the later archaeological material is transitional, forming the link between the pre-European and modern potter.[94]

In his further comments, Speck states (1928: 424–425):

> The modern Pamunkey have not quite left off making pipes. Some of the women . . . and some of the men . . . manufacture them as they were made two generations ago. They dig their clay in the same holes along the river. They gather and burn the mussel shells, and clean and mix the clay with the powdered shell in the same proportion, about one part of shell to five of clay. They burn them in the traditional way by piling a heap of dry fine sticks and a dozen or so dry cornstalks to the height of five or six inches, enough to cover two or three pipes which have been four or five days in the shade. Then when one covering of the sticks has been burnt off, the pipes are done and ready for use.

Pottery-making among the Pamunkey appears to have received a fresh impetus during the last decade. At the present time vessels and pipes are being manufactured primarily for tourist trade. The potters use—quite unsuccessfully, I think—painting and crustation (*i. e.*, pre-fired as well as post-fired applications of the color medium) for decoration; and they often employ a kiln for firing. That some aboriginal Southeastern methods still prevail at Pamunkey has been shown by the recent field work of Mr. Theodore Stern, a graduate student in anthropology at the University of Pennsylvania. Mr. Stern devoted several field trips in 1940 and 1941 to a study of pottery-making at Pamunkey, working chiefly with Paul Miles, the most active native potter. The results of his work have been embodied in an unpublished manuscript (1941)[95] placed in the Library of the University of Pennsylvania. The paragraph immediately following this is based on that manuscript and on various personal information given me by Mr. Stern.

At the present, two styles of pottery-making exist at Pamunkey; one follows native tradition, while the other is taught at the pottery school maintained on the reservation by the state since 1932. The traditional procedure may be outlined briefly as follows. Local clay, recognized as tribal property, is utilized, invariably tempered with calcined mussel shells in varying proportions up to one-third by volume. The constructional methods employ modeling, segmental procedure with morsels and with fillets, and molding. The morsel variant depends on amorphous increments of paste which are gradually added to the growing wall, starting with a previously modeled basal disk. The segmental fillet-building depends upon the circuit variant of annular procedure. (The principle is the same as that practiced at Catawba.) A double mold is used in the production of certain pipes, but more commonly pipes are modeled. The polishing pebble, especially valued and often quite old, is usually kept in active service through family inheritance. An impressive variety of decoration is generally executed; sundry impressions, rubbing, brushing, indenting, grooving, incising, etc., produce the desired effects. However, attempts at painting, presumably aiming to imitate Pueblo examples, have been, I think, totally unsuccessful. The firing of larger pieces is usually done at an open fire outdoors. Pipes are fired either in a metal receptacle or upon the grate of a kitchen stove. The traditional style has been flourishing for over a decade, and has retained its methods and improved its technique in most recent years. The school, at first adopting the circuit construction and using local clay, soon introduced radical innovations. Stern lists pot-molds, use of sand for temper, utilization of untempered clay, importation of factory clay, the potter's wheel, templates, paints, glazes, and the kiln. The influence on form is therefore pronounced, and reflects not only the new mechanical facilities, but also the pressure of tourist demand. The school promotes imitations of Pueblo shapes and decorations, and the first instructor is credited with the introduction of decorative pictographs purporting to represent Indian sign language. The school-taught style of pottery-making no longer retains any native Pamunkey methods, but it apparently does not usurp the local craft. At the same time its economic advantage is certainly of distinct importance.

Compared with the product of the Catawba (or its counterpart at Cherokee), the native Pamunkey ware is marked by a gross qualitative inferiority. There is historic proof and some extant evidence of certain influences brought in by Catawba potters marrying among the Pamun-

[94] At Catawba flat bottoms seem to be an aboriginal characteristic. However, Speck's relegation of this type of base in the Virginia-Maryland-Delaware coastal area to post-European contacts remains unchallenged.

[95] Mr. Stern obliged me with a copy of his manuscript, and gave me detailed verbal accounts of his several field trips. I hereby acknowledge his courtesy in allowing me to use his data in this paper.

key. In modern times, however, such instances have just about disappeared. The Pamunkey who still follow the native traditional methods obviously lack the skill, and perhaps also the zest, which is so characteristic of the Catawba potters. Their products seem devoid of any special care, finesse, and enthusiasm. While the techniques still preserve certain aboriginal principles, the finished products do not reveal them. Even an expert in native American Indian pottery-making, past or present, might well be perplexed by some of the modern "traditional" Pamunkey product. The tourist trade seems to be chiefly responsible for the current state of the Pamunkey craft. Despite the relatively low pecuniary gain —in contrast with Cherokee and Catawba—the economic aspect is stronger than the force of tradition. Nevertheless, the school should be able to work out a program which would serve to retain and improve the traditional style and at the same time to enable others to better their economic status through industrialized pottery-making. Such a program should avoid combining the two tendencies, for each has its specific scope, and each can satisfy the relevant demand.

After the foregoing brief survey of pottery-making at Pamunkey, it seems appropriate to inquire into the status of the craft among the neighboring remnants of Indian descendants in Virginia.

About 1925 Dr. Speck collected a plain, spheroid, fairly smooth, and undecorated jar of Chickahominy provenience, about four inches high and about five inches in diameter, recently made in the family of Chief O. W. Adkins.[96] The Chickahominy might still find it possible to demonstrate pottery-making reflecting some formerly used methods.

A similar supposition is probably also applicable to the Mattaponi and to the Adamstown or Upper Mattaponi. Only field investigations can determine the actual conditions; certain sporadic information seems to indicate that at least traditional knowledge of pottery-making may be expected.

According to Speck (1925: 69), "pipe making and ceramics, it seems, passed out of existence among the Rappahannock before the Civil War. . . . It is indeed unfortunate that some vestiges of clay-working did not continue until a later day, at least in memory, as they have among the Pamunkey and Mattaponi." For the Rappahannock, then, it would seem less possible to entertain such hopes as one may, albeit reservedly, for some of their consanguineous and culturally akin neighbors. And yet fresh field investigation of the question is desirable before final judgment is passed.

For the mixed descendants of the Namsamond, Potomac, Powhatan, and (?) Werewecomoco, now surviving in Virginia "in the same general locality where their ancestors lived" (Speck, 1925: vii), no positive records or traditions of pottery-making, past or present, have, as far as I am able to ascertain, been gathered.

In the Chesapeake region farther north, native pottery-making did not survive into modern times either in practice or in tradition.

In his monograph on the Nanticoke of Delaware, Speck (1914: 36–37), dealing with local archaeology, did not specifically mention the "smooth ware" simply because its general occurrence in the area was not then established. This, however, was done in subsequent years, and Speck (1928: 412, 424) was the first to report and to evaluate the culture-historic importance of such pottery.

More recently Davidson (1935: 6 ff.) reported the finds from the Slaughter Creek site, Delaware; his excavations brought forth, among other remains, undecorated sherds with crushed shell inclusions, *i. e.*, Speck's "smooth ware."[97] It is to be hoped that further systematic work in Delaware will follow, and that it may reveal evidence with which to reconstruct local chronology. Similar investigations are also desirable in Maryland. An attempt to establish the temporal position of the flat bottom, and above all the chronological relation of the "smooth ware," must necessarily wait for dependable archaeological evidence.

SOME HISTORICAL RETROSPECTS

In the deeper Southeast, pottery-making appears to have been abandoned during the nineteenth century[98] by the Chickasaw,[99] Chiti-

[96] Information from Dr. Speck. For an illustration of the specimen, *cf.* Speck (1928: fig. 113).

[97] Dr. Davidson informs me that he did not find any flat bottoms.

[98] Accurate dates are, quite naturally, extremely difficult to ascertain. The Caddo, not included here in this relation, may have made pottery as late as the nineteenth century. However, direct evidence seems to be absent in relevant documentary sources. Nor does one find any alleviating data in records devoted to the various known tribal components of the Caddo nation. Ford (1936: 72 ff.) deals with historic Caddo pottery; his dates go back to the seventeenth and eighteenth centuries. In a similar

macha,[100] Choctaw,[101] Koasati,[102] Natchez,[103] Pascagoula,[104] Seminole,[105] and Tunica.[106]. Details of technique and other aspects comparable to Catawba practices appear originally to have been possessed by the Chitimacha,[107] Choctaw,[108] and Natchez.[109] In the case of the Choctaw, moreover, there seems to be a parallel in the esoteric peculiarity of the imposed restriction regarding the witnessing of the firing process (Bushnell, 1909: 13). According to Holmes (1903: 143), "one specimen in the U. S. National Museum labeled 'Seminole' is identical with Catawba ware."

The technique of eighteenth century Louisiana potters, as presented by Du Pratz (1758, **1**: 124; **2**: 178 ff.) and by Butel-Dumont (1753, **1**: 154; **2**: 271 ff.), allows little comparison with the Catawba.[110] Notable exceptions exist in two points specifically brought out by Du Pratz (1758, **2**: 179): the use of a polishing pebble and its careful preservation during migrational movements; and the removal of constituent sand from the clay used for paste.[111] Butel-Dumont's account (1653, **2**: 271) of coiling with fillets "six or seven feet in length,"[112] seems to be without parallel in the Southeast. Speaking of Ewi Katalsta, who appears to have been the last native Cherokee potter to use coiling, Holmes (1903: 57) compared her true coiling method

manner Ford (1936: 40 ff., 72 ff., and 98 ff.) considers the case of the Choctaw, Natchez, and Tunica. Fairly accurate dates on the termination of the potter's craft among these three peoples are available; *vide infra*.

[99] Holmes (1903: 130), *i. e.*, "if the labeling of certain specimens in the National Museum is correct."

[100] Swanton (1911: 347): "Pottery continued to be made [according to the author's Chitimacha informant] until about eighty years ago." Especially significant is Swanton's remark: "The important part played by a monster pot in the [Chitimacha] flood legend indicates that the art was an old one among these people." The recollection of the absence of tempering material and of the exclusive employment of fingers on the interior of the vessel revealed by Swanton's informant is indeed interesting and valuable.

[101] *Cf.* Holmes (1903: 130), with the qualification regarding the National Museum labels; also p. 102. Swanton (1931: 40) states: "We have no information from early writers regarding their [Choctaw] pottery except the mere fact that they had it." Bushnell (1909: 13): "Pottery bowls are no longer made, although they are remembered by the living [Choctaw] Indians who recall having seen bowls provided with three small feet; consequently bowls must have been in use only a short time ago." The account of Butel-Dumont (1753, **1**: 154; **2**: 271 ff.), dealing with eighteenth century Louisiana potters, was considered by Mason (1911: 105) as applying to the Choctaw, without an explanation of the view; however, Swanton (1931: 62) regards it, with convincing deductions, as an observation made among the Natchez.

[102] Harrington (1908*a*: 406).

[103] The Natchez probably made pottery as late as the nineteenth century (*cf.* Swanton, 1911: 81). The account of Butel-Dumont (1753) and that of Du Pratz (1758, **1**: 124; **2**: 178–179) seem to be the earliest and at once the best available, even if deplorably incomplete, records on the subject. The use of color (Du Pratz, 1758, **1**: 124), presumably for monochrome painting, and the manufacture of figurines used in temples (Swanton, 1911: 159 ff.), bespeak an originally well-rooted pottery industry among the Natchez; Holmes (1903: 102) spoke specifically of the high status of the Natchez arts and industries.

[104] Swanton (1911: 303), quoting Margry: "They have plates made of wood and others of earthenware; they are all very well made, although by the hand of savages. The women of the savages also make large earthen pots, almost like big kettles, which hold perhaps forty pints, in which they have their hominy cooked for two or three families. . . . These pots are of clay (*terre grasse*) and of a round shape, almost like wind-mills." The form comparison has a multiple suggestive value; I venture to consider the possibility of the shape illustrated in Holmes (1903: pl. LVI*a*—specimen from Bear Point mound, Alabama, p. 106), which, however, resembles the body of a modern churn as well.

[105] Holmes (1903: 130), *i. e.*, "if the labeling of certain specimens in the National Museum is correct."

[106] Swanton (1911: 315, 319), quoting Gravier: "Earthenware pots, quite well made, especially little glazed [?] pitchers, as neat as you would see in France." "Little earthen pots . . . used for religious purposes."

[107] *E. g.*, the "peculiar pipe, into which a number of stems could be inserted" (Swanton (1911: 349)). This is strongly suggestive of the Catawba "peace pipe."

[108] *E. g.*, the three-footed vessel, the use of grease on freshly fired specimens, the absence of temper in paste for pipes, the incised decoration, and the lustrous finish (*cf.* Bushnell, 1909: 12–13).

[109] *E. g.*, greasing of vessels after firing (personal information from Dr. Swanton). The sporadic cases of burnt, crushed bone tempering medium, noted in a few sherds collected at Catawba, must be viewed with reserve.

[110] Holmes (1903: 57, 102) qualified both of these instances as "inadequately described," yet considered them to be "brief but valuable records of the practice of the art in this section." "But," he added, "we are not definitely informed which of the various people were referred to in their accounts." As already pointed out, Swanton assigned the two sources to the Natchez, while Mason considered Butel-Dumont's description as applicable to the Choctaw. Extremely interesting and important, I think, is Butel-Dumont's record (1753, **2**: 271 ff.) of true coiling, which, as far as I am aware, is the historically first notation of this method in the Southeast.

[111] It is of further interest to note the reference to the use of a "*bois plat*" upon which the women worked. The flat wooden surface may, quite conceivably, have had some effect upon the plane of the bottom, possibly a flattening one. On this point Du Pratz is silent; it seems well, however, to recall the flat base in ancient Catawba pottery.

[112] A single fillet of such length would be very difficult to prepare and to manipulate. Perhaps the author had in mind the computed measurement of several fillets used in the construction of a given vessel.

with that of the "potters of Louisiana," but gave no tribal names. One is led to the inference that Holmes (1903: 57) here relied on the work of Butel-Dumont, which he excerpted at length in translation.[113]

For the modern Yuchi in Oklahoma, formerly of the Southeast, there is Speck's ethnological study (1909) from which the following quotations on pottery-making are germane to the present purpose: "The clay is washed to reject grit . . . the lengths or sticks of rolled clay are coiled around on . . . the . . . base and so built up until the proper height and form is obtained" (pp. 25 ff.).[114] The scraping with the edge of a mussel shell, rubbing and polishing, the manner of progressive firing (with frequent turning), the rejection of the grit, the preparation of the base disk, and the exclusively female operators (in the manufacture of vessels) (pp. 25 ff.) parallel the Catawba practices. According to Speck (p. 28), the modern Yuchi vessels resemble more the Chesapeake-Potomac group than "they do the highly ornamented and complex forms of the Southern Appalachian groups." The pipes, still made by men, and of the same paste as the pots, are hand modeled; plain and effigy forms are common (pp. 28 ff.). Here too, then, in certain points of technique, and in the plain shapes, a comparison with Catawba pipes can be drawn.[115]

CONCLUDING REMARKS AND COMMENTS

With the Catawba, pottery-making not only survives along traditional lines, but is quite commonly practiced. With the exception of changes in the form, modern influences are negligible[116] and readily distinguishable. The most recent accomplishments are permeated with the same elements of techniques which were current at the reservation in the eighties.[117] These were rooted then—although minimized by Holmes (1903: 143)—as they are now, in aboriginal antecedents. It must be stressed again that Holmes wrote about the Catawba potters on the basis of Palmer's and Mooney's records, and from examination of museum specimens, rather than from actual field observation. That probably explains his failure to grasp the technique in the light of then existing conditions and retrospective values.

An impressive continuity of basically native forms amid modern Catawba ware is demonstrable from extant collections, and is also supported by certain archaeological comparisons. I realize, of course, that the "ancient" evidence is not adequately datable to permit far-reaching retrospective reconstruction of a chronological value. The inherent limitations, in view of the existing deficiencies, are obvious. My deductions in this respect, therefore, are not necessarily applicable to the pre-Columbian time scale. Nevertheless, it is significant to note that the various samples of archaeological material found in sites within the Catawba territory show germane similarities to the modern product, and that they apply alike to form, textural quality, surfacing, firing, and constructional technique (*i. e.*, annular building).

It remains for future archaeological investigation to ascertain the history of Catawba pottery-making as an integral part of their material culture. So far as the past is concerned, it is evident that ample material remains exist in the hitherto very little explored sites in the Catawba locus. Naturally, an archaeological reconstruction must arise from such tangible evidence as may justify retrospective interpretations. Although pottery in itself is not adequate, nor exclusively dependable for this task, it seems reasonably presumable that a study of the existing collections, and especially fresh field work, should establish a comprehensive index with which an objective plan may be formulated.

It is well known that at the present time the archaeological history of the Catawba remains virtually untouched. Yet it is clear that the contemporary pottery of this people discloses

[113] He also gave a similar treatment of Du Pratz's publication.

[114] I am obliged to Dr. Speck for the following additional information. The rolls were of irregular length, yet applied in the manner of the Catawba circuit procedure. The Yuchi were no longer making pots by 1908, *i. e.*, the date of his study here quoted.

[115] It is interesting to note the survival tendency in pipe-making in contrast to that of pots. This already has been shown to be true of the Pamunkey and the Choctaw.

[116] *Cf.* Holmes (1903: 143): "But an examination of numerous ancient sites and a number of mounds in the region occupied by the Catawba in early historic times, yields forms of vessels distinctly western in some of their features [this is *à propos* of Western and Eastern Siouan comparisons], and in cases there appear also pretty well-defined characteristics of the historic Catawba work. . . . Specimens found on the older dwelling sites of the people resemble the modern pottery in color and finish, but they are of better workmanship, and the shapes resemble less closely those of the whites. All are flat-bottomed, have the thick walls and peculiar color and polish of modern Catawba ware, and are well within the Catawba habitat, even if not from sites inhabited by them in historic times."

[117] *Cf.* Palmer's observation of 1884, incorporated in Holmes (1903: 55), and those of Mooney, also published therein (*ibid.*: 53 ff.).

many characteristics of the ware gathered at sites which are attributable, either traditionally or through sources, to former Catawba occupation.[118] The manufacturing techniques of the two categories of ware are almost identical. Analyses of the history of yet other elements characteristic of the cultural attainments of the Catawba seem plausible of accomplishment. They should aid in the retrospective tracing of the last remnant of the Eastern Siouan-speaking group. The recent works of Speck (1935) and Swanton (1923, 1935, 1936)[119] reveal new evidence of intraregional as well as extraterritorial distribution of Eastern Siouan peoples in aboriginal times.[120] Toponymy and other linguistic traces indicate their former existence quite far afield from the northern section of South Carolina. Archaeologically, there is as yet very little in the way of adequate light with which to aid in these endeavors in so far as the Catawba are concerned.

The recent archaeological explorations in the Southeast have considerably advanced the understanding of culture historical events in areas lying to the south of the modern Catawba habitat. At the same time, however, a comprehensive regional system of chronology is yet to be formulated.[121] In North Carolina systematic archaeological research began in 1936, while in South Carolina no such effort has yet been made. Turning northward from the Catawba locality, one meets first with a pronounced archaeological lacuna in inland Virginia, where pottery remains have received scanty attention thus far. However, in Tidewater Virginia, at least from the Chickahominy River, and in an area extending northward into Maryland and Delaware, there is the "smooth ware" described by Speck (1928: 399 ff. and especially 412). Its fabric, surface finish, and color tally with the comparable examples from the Catawba and Cherokee areas.

In South Carolina, I feel, the logical and perhaps also most promising start toward an understanding of Catawba archaeology should be with the known sites situated in the country of the contemporary, albeit ethnically and culturally changed, congeners. Old collections necessitate careful examination, but, above all, more field work is imperative. The results thus far attained by the University of North Carolina explorations certainly inspire hopeful expectations. The surviving pottery technique and the continuity which so prominently mark the craft of the Catawba should provide valuable aid in future inquiries. The usefulness of pottery in attempts leading to reconstructive studies of culture history is a stimulating factor in the case of the Catawba. However, the investigator will find other material remains equally helpful in the preparation of a repertory (in so far as extant evidence may suffice and permit), with which to comprehend events of the past. Subsequent procedure should apply such data in cognizance of those leads already established by Mooney, Speck, Swanton, and others.

The dangers of entertaining extraterritorial comparisons before the local developments are adequately reconstructed certainly need no elaboration. Yet it seems in point to refer to some of the similarities between modern Catawba and modern Pueblo processes of manufacture. The latter are well known from Guthe's splendid study (1925). In 1929 I sent some Catawba specimens to the Museum of Anthropology at the University of Michigan and supplied a brief description of the technique. Dr. Guthe's letter to me, dated March 8, 1929, confirmed several similarities. With Dr. Guthe's permission, I am now able to quote his comment as rewritten by him for this purpose:

> The similarity of Catawban pottery making to the method used in the Southwest is striking. The steps you have outlined are practically duplicated in the Pueblo area. In building a vessel the strips of clay to be added are shaped as ropes and do not form rings until they have been applied to the vessel. In the Southwest there are usually two additional steps, namely slipping and painting, between your seventh [drying] and eighth [firing] steps. Instead of using the domestic fireplace as the course of the heat, the Pueblo people build an oven of dung cakes, which are used as fuel, and the firing is done out of doors. The irregular black spots would be explained in the Southwest as a result of contact between the vessel and

[118] *Cf.* Speck (1928: 413). As previously stated, the University of North Carolina archaeological survey has recently identified several historic Catawba sites. It was my privilege to examine the material deposited at Chapel Hill through the courtesy of Mr. Joffre L. Coe. The sherds show close resemblances to the modern ware in forms, surface finish, color, the *flat* bottom, and constructional technique.

[119] Also Dr. Swanton's personal information; his paper presented at the Annual Meeting of the American Anthropological Association, Andover, Massachusetts, December 29, 1935; and his lecture before the Archaeological Society of New Jersey in Trenton, New Jersey, January 18, 1936.

[120] *Cf.* also Bushnell (1934, especially pp. 8–9, and map 4).

[121] Ford and Willey (1941) attempted an interpretative synthesis which clearly indicates the many existing gaps.

some object, such as dung or wood, which would cause a deposit of carbon. The Pueblo potters secure their beautiful black ware by smothering the oven and the vessels within it immediately after firing, with a quantity of new, pulverized fuel obtained by crushing dung cakes. This process is closely similar to that which you mention. The method of inserting handles is the same as that used in making horizontal loop handles in the Southwest. I do not know of any cases of moulds being used among the Pueblo Indians.

Pottery-making at Catawba is demonstrably an old cultural heritage propagated by a conscious evaluation of its economic expedience and importance. While its domestic utilitarian function has diminished with the substitution of modern metal utensils, the ware is produced not merely for the sake of profit. The element of the potter's satisfaction with her skill and the aesthetic feeling derived from her endeavors are likewise prominent factors. The artisan certainly taxes her individuality, and a meticulous care governs the manufacturing process. Patience, dexterity, prudence, and gratification are responsible for the final product which reflects artistic as well as utilitarian values.

It is evident that the individual potter knows little or nothing of the principles of chemistry and physics which inevitably play important roles in the various manufacturing processes. Yet it is equally true that some of the properties and changes which may be attributed to such principles are recognized and respected through the teaching of experience. Questioning on the part of the inquiring observer may not always bring satisfactory answers; the artisan, however obliging, often wonders why "obvious" matters should call for elucidation. Whether the final product represents a sophisticated vase form, a set of bookends, or an ash tray, its appearance and quality are of the same nature as the general run of the purely native-inspired pieces. The color, lustre, and the mottled effect, which render the Catawba ware so characteristically individual and recognizable, are constantly present.

It seems, then, that despite wholly artificial outside forces, which have been especially prominent since contact with the whites, the potter's craft of the Catawba is being conservatively perpetuated [122] in consequence of the two basic factors responsible for its existence: the supply and quality of the necessary clays, and the custom-taught technique. So long as these two conditions persist at the reservation, the continuity of the craft as it now exists should prevail. The craft is now rather static, although quite lively; a similar status was noted over fifty years ago.[123] Certainly, the old masters, whose pride and dexterity are respected by their pupils, show an ardent desire to preserve the standards by imparting their knowledge and experience to the younger generation. Catawba pottery-making, a common possession of the "nation," illustrates an interesting case of cultural continuity which demonstrates the role of a direct traditional survival. Furthermore, Catawba potters married among the Cherokee[124] appear to have consistently preserved their own techniques. And they instructed local talent in such manner that the contemporary Cherokee craft is now dominated by Catawba practices.[125]

It will be recalled that Holmes (1903: 143) depended largely on the criterion of form in so far as he concerned himself with historical retrospects of Catawba pottery. Yet, as already quoted, he recognized certain comparisons of the product of the historic Catawba in archaeological sites within their early historic habitat. It is difficult, therefore, to appreciate his comment (1903: 143): "The modern Catawban pottery has been so modified by post-Columbian conditions that few of the original characteristics are left, and comparison is fruitless." I have stressed the point that modern commercial influences appear to have affected only the *form* of Catawba pottery, and that "ancient" forms are still currently produced. With respect to technique, no appreciable differences exist between the two categories of ware. Unfortunately, the designa-

[122] Mooney (1894: 74), referring to the Catawba reservation population in the year 1889, wrote: "The women still retain their old reputation as expert potters."

[123] Palmer visited the Catawba in 1884, at which time he took notes on their pottery-making and later tendered them to the National Museum (*cf.* Holmes, 1903: 55). Historically, this appears to be the first record of the Catawba craft. Holmes's account contains excerpts from Palmer's work; the technical details tally with my observations, but there are no interpretative comments. I have not examined Dr. Palmer's original field notes to be able to state their full nature. The notes, according to Holmes, should be in the National Museum; my attempts to consult them in Washington were unsuccessful.

[124] First noted by Mooney (1900: 165) and Holmes (1903: 53). (Not mentioned by Harrington.) According to Speck (1913: 330), "half dozen or so persons of Catawba blood . . . were, in 1913 . . . living and mixed with the Cherokee." Two of these were women practicing pottery-making in the Catawba tradition (information from Dr. Speck).

[125] *Cf.*, in this connection, the cases of Susan Owl and Ewi Katalsta (*supra*).

tion "ancient" has no satisfactory connotation, in as much as the pottery of the Catawba is still unsubstantiated in point of chronology. Nonetheless, it seems significant that an impressive degree of continuity is definitely suggested by the known meager, but positive, evidence.

The cultural and temporal position of Catawba pottery-making offers an important and promising subject of research. The chief needs toward further appreciation, understanding, and interpretation of the subject, seem to be the following:

1. Elucidation of certain phenomena of the firing process whereby certain mottled effects are attained.
2. Investigation of the question of historical priority of the ring variant as against the circuit variant of annular construction.
3. Study of all material remains from former Catawba sites.
4. Formulation of a comprehensive index arising from (3), with which to reconstruct, so far as is possible, the totality of the former cultural expression of the Catawba.
5. Reconstruction of chronological sequence and formulation of analogies with adjacent regions in order to establish relative dating.

The technicalities presented here are based on tangible evidence; and the interpretative data rest on observations and inquiries gathered in the field. If I have failed, in this brief presentation, to justify every reasoning, the fault may well inhere in my incomplete understanding of certain aspects of Catawba pottery-making. To deal with the contemporary craft is simple enough, since recourse to inductive means is available. So far as past events are concerned, the difficulties increase the more one projects one's aims into antiquity. Examination of literary sources dealing with the Catawba reveals a striking similarity with the contemporary fundamentals underlying the individual processes. The ring building variant, however, does stand out as a signal exception; but it by no means appears to be a very recent phenomenon.

Purely archaeological considerations of Catawba pottery-making suffer from serious shortcomings, among which the lack of appropriate exploration and adequate appraisal of extant remains are particularly acute. Consequently, retrospective contentions must provisionally rest on rather uncertain premises; yet such knowledge as can be applied with a reasonable margin of assurance proffers at least some stimulating support.

The mosaic which I have here attempted to conjure up is certainly neither complete nor adequately balanced. Its very foundation is distorted owing to the pronounced lack of uniformity in historical perspective, and the individual components plainly reflect unequal values. The endeavors and results here described are tantamount to no more than a mere scanning of the horizon, and the potential opportunities are certainly not exhausted. Yet it is to be hoped that this paper may be of some service in further investigations in the northern periphery of the Southeast. I firmly believe that Catawba pottery-making, recent and ancient, furnishes a substantial part-basis for a point of departure toward a broader appreciation, understanding, and interpretation of the culture history of this important area.

ADDENDUM: COILING

The motivation for this note on coiling inhered in a dilemma which faced me in considering the choice of a suitable term for the Catawba ring variant of annular construction. In my early field notes, and for some time thereafter, I called the process "pseudo-coiling." My search in literary sources dealing with the North American field failed to reveal a parallel phenomenon;[126] but I did notice that "ring building" was equated with "coiling." That led me to investigate "coiling" more in detail, and I noted yet other definitions such as "moulding," "strip building," "adding circles," "pressing or coiling process," etc. Eventually I decided to call the aforementioned Catawba variant exactly what it is, *i. e.*, ring building, and to submit the so-called "coiling" to a scrutiny. The result is the present discourse, and its purpose is to reveal the laxity in description (if not in concept) of the process, to indicate the untenable looseness in terminology, and to outline a classificational grouping of the various methods of fillet construction.

Of the several variants of pottery construction, the process called "coiling" is most frequently found—*in literature*. Whether directly stipulated as such, or merely inferred, the term is rarely applied without ambiguity. Obviously, coiling denotes a building process and must not be confused with other steps of pottery manu-

[126] Thompson (1930: 95) reports the occurrence of true ring building among the modern Maya potters at San Antonio, British Honduras.

facture, even if its function be an aid in such aspects as form-giving (shaping) or embellishment. Sources on pottery technique are quite devoid of a standard definition of the phenomenon and its distinguishing criteria. Although "coiling" is often explicitly contrasted with other means of constructional procedure, considerable confusion exists in terminology and in classificatory segregation. It cannot be said that the term "coiling" connotes a recognized inclusive designation, *i. e.*, that it embraces all varieties of pottery construction which depend on the use of paste in the form of a fillet ("rope," "rod," "strip," "coil"). On the contrary, the customary reference to "coiling" lends itself, as a rule, to one of two interpretations:

1. Continuous operation with a single fillet or with several individual fillets, successively connected and forming a spiral.
2. Construction with the use of several fillets, progressively applied, and individually bonded, but not successively linked.

These modes comprise perhaps the most commonly recognized variations. They are, however, rooted in quite different principles. The first depends on carrying the paste medium around in the manner of a superimposed spiral regardless of the actual number of fillets employed. The other involves concentrically placed individual fillets; this is illustrated by the circuit and ring variants of the Catawba.

Lexically, the word coiling fails to provide a satisfactory denomination. Its etymology is quite unclear, its meaning ambiguous; as injected into pottery terminology, it reflects ill-chosen borrowing. Dissatisfaction with so loose a term has often been voiced or implied; yet the catachresis continues. In view of the circumstance that the word "coiling" is so thoroughly implanted in American literature, and especially by reason of its application to prehistoric Pueblo pottery by Holmes (1886: 257 ff.), it would seem futile to consider a substitute. However unpalatable the misnomer may seem, it does carry a connotation which is objectively measurable; its chief criteria can be stipulated with precision and clearly defined. It may be useful to choose a collective term under which to include coiling, circuit procedure, and ring procedure; *fillet building* seems to serve the purpose.

For the present needs I am guided by the following understanding of coiling in the potter's craft: Coiling implies a building process which is accomplished either with a single fillet (apparently a rare phenomenon) or with a series of fillets successively connected to constitute a continuous chain. In either case the paste medium is fashioned first, and its application depends on a purely manual manipulation in which the potter's hands must be free to effect the placement and bonding. The fillet, single or chained, is carried on spirally, and in such manner that each loop (irrespective of the shape of its course) is equivalent to the circumference of the wall under construction at a given height. However, the fillet itself does not terminate as it completes the circumference, rather it is carried on into the subsequent volution. Owing to the unavoidable overlapping which results from the continuous course of the fillet (single or chained), the plane of a given loop is necessarily an uneven one and often slanted. Whether the construction begins with or without a previously fashioned base (bottom), and whether the fillets are completely obliterated or not, coiling inevitably involves unbroken linkage in carrying the paste medium around on a spiral course. When the operator begins by coiling the base (bottom) first, that process entails convolute coiling; the fillet is then carried around plano-spirally, and the manipulation constructs a more or less flattened disk. A similar base (bottom) may be procured by modeling, *i. e.*, by direct shaping in hands (combined, perhaps, with an impact on a planed, solid surface); or it may be achieved by pressing the paste within or upon a shape-giving contrivance (sherd, pot, basket, weighted bag, etc.), *i. e.*, a pseudo-mold (as distinguished from a true mold within which or upon which to produce a vessel, figurine, appliqué ornament, pipe, lug, handle, etc.) With the convolute, base-producing variant, the wall-erecting process continues without interruption. On the other hand, a previously modeled base

FIG. 32. Over-all coiling (left); coiling the sides upon a flat base (center); and vessel with coiling completed (right).

FIG. 33. Dried and fired pieces, broken to show methods of construction.

(bottom) delays coiling until the wall itself is begun. The fillet (single or chained) is then carried in the manner of a vertically superimposed (or perhaps juxtaposed) spiral. For, having completed the course necessary for fully describing the requisite circumference, the fillet is not terminated (by bonding its margin, or by breaking off the surplus length), but is retained as an integral part of the subsequent loop (or loops). In other words, the fillet (single or chained) is retained beyond the length equivalent to the circumference of the wall, and the potter proceeds with construction by carrying, from volution to volution, the appropriately fashioned medium of paste throughout the coiling operation. When convolute procedure is turned into a spirally wound course, a case of over-all coiling obtains; otherwise a combination of coiling and another method (or methods) is at play. Imbrication, although common, is not necessarily an index characteristic of coiling. Coiling leads to the construction of either a blank (subject to separate shaping) or an essentially final form. In the former case the shaping of the desired form becomes an entirely separate task which is, of course, independent of the constructional steps. The form-giving manipulation, accomplished either purely manually or with tools (anvil, paddle, spatula, spoon, trowel, etc.), is fully detached from coiling. (It is inconceivable to "*coil* with an anvil, paddle, stick," etc.) It is plausible, of course, to produce embellishment concurrently with coiling. This may consist of seams, overlaps, protuberances, indentions, corrugation. (Corrugation, a specialized tecto-decorative process, logically necessitates concurrent form-giving.)

Viewed in the tecto-functional sense of its implications, coiling proper is distinguished by an ample number of characteristics. Naturally, these characteristics are not constant; nor need they always occur in like combination. However, those which invariably do distinguish a case of coiling as such, are sufficiently well pronounced to warrant—indeed to compel—differentiation from other fillet-using processes. And, quite conceivably, a potter may, in accordance with certain conditions, combine two or more modes of construction in the course of erecting a single vessel.

In the Southwest, it is commonly alleged that Pueblo pottery was made by "coiling," while the Hohokam ware, *in contrast*, was made by "the paddle and anvil method" (*cf.* V. J. Fewkes, 1941*b*). Roberts' initial attempt (1935: 20) to dispel this popular confusion appears to have gone unheeded, perhaps chiefly because of the casual manner in which it was voiced.[127] Gifford (1935) obviously missed his opportunity in commenting on the praiseworthy effort of Roberts, by failing to elaborate the fundamentally different functions of "coiling" on the one hand and of "the paddle and anvil method" on the other hand.[128] However, Roberts appropriately reiterated his warning when, comparing Hohokam and Anasazi wares, he stated (1937: 20):

> The pottery made by the two patterns differed in certain respects. Both groups used the coiling method but the finishing process varied. The Anasazi smoothed the surface of their pots with scrapers and polishing stones; the Hohokam completed theirs by employing a paddle and anvil.

"Coiling," or rather *fillet building*, as applicable to pottery, connotes a constructional process which, by virtue of logic as well as of physical law, unconditionally requires manual manipulation. The potter's hands simply must be free for the handling, placement, adjustment, and bonding of the fillets. The use of the paddle and anvil leads to deliberate alteration of either a partially or a completely built, fully plastic, embryonic vessel. Such alteration may involve: thinning of wall; compacting of paste; bonding of tectonic segments; obliteration of junctions;

[127] *Cf.* also Roberts (1936: 527) for a somewhat revised version.

[128] I shall have occasion to refer later to Gifford's paper of 1928, which deals with the "two methods" in question. Yet it seems well, in this connection, to recall that he recognized (1928: 353) "two methods of *making coiled* pottery . . . in the Southwest." "The principal criterion of method," continued Gifford, "is the use or non-use of a wooden paddle and a stone or pottery anvil in *shaping* the vessel." (Italics in these quotations are by the present writer.) The fallacy is self-evident; it was simply restated by Gifford (1935).

smoothing of surface; shaping of final form.[129] However, no matter what the ultimate effect may be, the paddle and anvil, either in their dual role or independently, cannot, again by virtue of logic and of physical law, be utilized as an aid to the constructional process proper. Yet it is conceivable that bonding and adjustment of fillets might, if desired, be accomplished with the use of a paddle, stick, anvil, etc., although these processes are most commonly achieved by plain hands. In any event, the manipulation of the paste segment, as it enters the building process, indispensably requires that the potter's hands be free. And once the paste becomes a tectonic component of the growing vessel, its subsequent alteration, no matter how attained, no longer constitutes construction. It is, therefore, a methodological error either to compare or to contrast "coiling" with the use of "the paddle and anvil," for the two are rooted in different technological principles (*cf.* V. J. Fewkes, 1938: 197). Perhaps the circumstance that considerable laxity has existed in constructional terminology contributed to the general confusion just considered; but the fact cannot be minimized that authors have often failed to be explicit in presenting their data. (*Vide infra.*)

A limited number of citations, all but one applying to the Americas, is here assembled as documentary evidence. The comments must not be construed as aiming at criticism, but rather as a means toward illustrating the urgent need—as I see it—of precise distinctions. I may add that I am fully aware of the limitations inherent in an attempt to reconstruct pottery technique from fragmentary fabrics (*i. e.*, fired pieces); naturally, indubitable criteria can best be illustrated from extant practices of pottery-making. Moreover, *I am concerned primarily with the fillet-using processes;* for this reason I shall mention other means of construction only in support of the distinction between the technicalities falling within the scope of my task. It becomes necessary, of course, to extend this inquiry beyond the limits of the Southeast.[130] Admittedly, no pretense of an exhaustive treatment is here maintained. Although I have deliberately restricted my inquiries to North America (with very few references to Middle and South America), the present discourse aims at a sampling rather than at a thorough treatise. Such aspects as time and space, historical perspective and correlation, as well as specific cultural considerations, are quite outside the scope of the task at hand. My endeavors at the moment are centered on a description of technicalities revealing sundry criteria of various modes of construction depending on fillets.

It is convenient at this point to deal with the selected citations.[131]

Wissler (1938: 69):

> As a rule all the New World potters used the coil method, *i. e.*, slender rods of clay were rolled out to convenient lengths and the vessel built up spirally. In some vessels from the Pueblo area the original traces of the coils were retained as decorative motives, but, as a rule, the surfaces were afterwards scraped smooth and to the required thinness.

---

[129] Incidentally, the various possibilities of using the paddle and anvil render the alleged "diagnostic marks of the method" of doubtful value indeed.

[130] In the Southeast apparently only the Natchez, as observed by Butel-Dumont (1753, **2**: 271 ff.) (tribal identification after Swanton, 1911), employed the single fillet (alleged length six to seven feet) process, beginning, it seems, with coiling the base (bottom) first.

The Cherokee, as has been shown, practiced coiling by linked fillets, but proceeded upon a previously modeled base-disk. Coiling in ancient Cherokee ware was inferred by Holmes (1903: 163).

During the excavations of the Etowah mounds "bits of coil," "discarded ends," and an example of a "rim finish with a coil" are said to have been found in the deposits (Ashley, 1932: 122 ff.). Ashley also stated (p. 109) concerning the mode of manufacture: "This method [that of the predecessors of the Cherokee], the one most commonly used in the Southeast, was coiling—the building of the side by placing strips of clay in spiral form, a base of the same clay being used as a foundation or nucleus." (Evidently, the linked coiling variant is implied.)

According to Claflin (1931: 19), "There is unmistakable evidence on several of the textile-marked sherds [Stalling's Island Mound, Georgia] that the coiling method of manufacture was in use." (Claflin did not present any elucidation as to the nature of the evidence, nor the manner of construction.)

Among the hitherto unpublished archaeological pottery from the Southeast in the Peabody Museum, Harvard University, I noted repeated examples of coiling by linked fillets, but none by one fillet alone. The specimens were from Georgia, Florida, Alabama, Louisiana, and Tennessee.

Most of the cinerary urns from two burial mounds at Cox, near Darien, Georgia, now in the Irene Mound Collection at Savannah, display ample evidence of all-over coiling by linked fillets. (Unpublished.)

The Southeast is here understood in the sense of the geographical definition given in Speck (1907: 289).

[131] The order in which these citations appear has no particular bearing on the subject. Indeed, it would probably be difficult to arrange a "logical" progression. It will become apparent, however, that some grouping has been reached, in as much as works of a general nature are considered first; sources on the Southwest are arranged in a chronological order; and interpretative compilations appear last.

The "convenient lengths" or "rods" indubitably signify coiling by linkage of individual fillets; and the spirally carried building procedure leaves no doubt as to the principle involved in the construction.[132]

Linné did not distinguish circuit building from true coiling. To quote from his well-known work (1925: 78–80):

> The [coiling] process has too often been described to call for any closer explanation here. In America the vessels are built from the bottom upward, whilst here and there, in other places, the building is commenced from the mouth portion. Frequently cracks occur in the joints between the coils, and the presence of horizontal cracks in primitive pottery is surely satisfactory proof of the technique used.[133] . . . The coils are usually the length of the vessel's circumference, each "story" consisting of one roll of clay.[134] At Cobreas . . . a woman potter made each "story" in three sections, joining each to its neighbor and to a clay roll underneath. The Guató, on the contrary, built up their vessels of a continuous coil, spirally laid on.[135] . . . It has been mentioned in the foregoing that the coiling method obtains, or has obtained, over the whole American continent.[136] . . . Terminologically, the word "Spiralwulsttechnik" would not seem particularly well chosen, as, judging from the description given, there is only one tribe, the Guató, who builds in a continuous spiral.[137] . . . The others, as has been described in the foregoing, by level courses of coils. The method [*i. e.*, true coiling] may possibly have been used in Porto Rico.

And commenting on Schmidt, with reference to pottery technique in the high culture area of the Andes, Linné held (1925: 80–81):

> There is hardly any alternative: here [in the Inca region, archaeologically speaking] the technique *must have been* that of coiling.[138] . . . As has been said before, from time of the discovery of Peru no description exists as to pottery making. Nor have I seen vessels published that could be adduced in proof of the presence of the coiling method. In a clay vessel from Huamachuco . . . the building-up coils seem, however, possibly to have been left unfinished.[139] But Wissler . . . says that it [coiling] was used by the side of shaping by means of a mould, and the same is stated in the case of Ancon. . . . Prof. Uhle, when questioned, stated his *absolute opinion* [*sic!*] that *all* archaeological Peruvian clay vessels, other than those of the Chimu culture, were made by the coiling method.[140]

The inadequacy involved is self-evident.

In distinct contrast to Linné's criticism (1925: 79) of the German term "Spiralwulsttechnik," I consider this nomenclature by far more appropriate and expressive than the English usage of *coiling*. Selecting a remote geographical example for the purpose of illustrating the usefulness of the descriptive value of its connotation, I quote Schurig (1930: 65): "Die Wulsttechnik, baut aus dünnen runden Lehmwülsten, . . . spiralig gewunden die Gefässe auf." Here, it seems to me, there is no ground for speculation; the spirally carried manner of building, depending on thin fillets, unquestionably linked as the spiral winding signifies, is explicit.

Harrison (1928: 31–32):

> Of the building methods . . . one of the processes has not only a striking individuality of its own, but has a wide and interesting distribution. This is the coiling method, especially characteristic of America, but also occurring in . . . widely scattered regions. . . . The characteristic feature is the preliminary preparation of rolls of clay. . . . In the process of shaping the pots *these rolls are coiled spirally upon each other*,[141] *the diameter of the spiral—or sometimes the circles*[142]—increasing or diminishing to suit the varying width of the pot[143] at its different levels . . .

[132] With respect to Wissler's contention that all the Amazon tribes had the coiling method, *cf.* the following quotation from Linné.

[133] Strictly horizontal "cracks" indicate parallel planes of individual annular fillets, whereas in true coiled pieces a slanted plane is usual.

[134] This is again proof of circuit building, rather than true coiling.

[135] Evidently Linné was not without examples of true coiling.

[136] This is a sweeping statement—in view of the ambiguity of Linné's understanding of the method.

[137] Farabee (1922: 86 ff.) implied linked coiling for the Conebo and the Sipibo.

[138] *I. e.*, either true coiling, or annular building? (Italics are those of the present writer.)

[139] Perhaps in a manner similar to that shown in Linné (1925: fig. 10).

[140] *Cf.* the warning of Shepard (1936: 440). (Italics are those of the present writer.)

[141] *I. e.*, true coiling, accomplished, in this case, by the linkage of several "rolls."

[142] Presumably, such "circles" form individually closed "rolls," each separately terminated upon a horizontal plane either by previously cutting the rolls to a uniform length, or by pinching the ends off as the circuit is completed; whether the ends of these rolls are joined by overlapping, or by simply connecting the meeting termini, the resulting bonding does not disrupt the plane continuity of the horizontally laid circuit. This, then, certainly is not true coiling. (Italics are those of the present writer.)

[143] *I. e.*, in the case of the form being shaped, provisionally or definitely, simultaneously with the building of the wall; otherwise a cylinder may be built first, its form subsequently altered, and the final shape brought out by a separate manipulation. *Cf.* the practice of the Catawba

it is very rarely that the finished pot shows any signs of its spiral or circular origin;[144] nevertheless, some of the old wares of the Pueblo and other Indians [?] of North America, and the Choroti[145] and others of South America, have the coils left in evidence as the basis of the decoration.[146]

An inconsistency is again demonstrated.

"Coiling" is invariably associated with Pueblo pottery-making, either explicitly or implicitly. The following quotations illustrate this.

Holmes (1886: 273–275):

> The ancient Pueblo potter rolled out long, slender fillets or ropes of clay, varying in width and thickness to suit the size and character of the vessel to be constructed. . . . When they were properly trimmed and smoothed, the potter began by taking the end of a single strip between his fingers, and proceeded to *coil upon itself*, gradually forming a disk. At first the fillets overlapped only a little, but as the disk grew large and was rounded upward to form the body of the vessel, the imbrication became more pronounced. The fillet was placed obliquely . . . and was exposed on the exterior side to probably one half of its width. *Strip after strip was added, the ends being joined so that the continuity might not be broken until the vessel was completed.*[147]

Holmes (1903: 372) referred to this as the "process known as 'coiling',"[148] and remarked that "the support [*i. e.*, the base-retaining means, or the puki; *cf.* Guthe, 1925: 31] was not a mold in the ordinary sense." *This account of Holmes may well be considered a classic description of "ancient Pueblo" coiling.*

Cushing (1886: 489) described the process of a Zuni potter thus: "She coiled around and around a center to form the bottom, then spirally upon itself, now widening the diameter of each coil more and more, then contracting as she progressed upward until the desired height and form were attained." There is no specific indication as to whether one or several fillets of paste were used.

However, J. Stevenson (1883: 375, 329), also writing of the Zuni, spoke of "additions of strips of clay," and expressly said that "traces of the addition of each strip . . . were removed before another . . . strip was added." This left some doubt as to whether circuit building or linked coiling was meant, although the use of several strips of paste was clearly established. The uncertainty, however, may be said to have been removed by the statement of Mrs. M. C. Stevenson (1904: 375): "The vessel is then formed by the successive addition of strips of paste long enough to encircle the bowl." This, unquestionably, identifies the process as circuit building.

J. W. Fewkes (1909: 53), by his reference to "corrugated or coiled ware," a practice with which one commonly meets in American literature,[149] implied linked fillets.

Kidder (1915: 414), speaking in broad terms, said: "The upper parts of ollas were constructed by the regular coiling method." Evidently, true coiling was meant, without, however, specifying whether any one or all possible variants of the methods were intended to be signified.

Nelson (1916: 168) used the designation "Corrugated or Coiled Ware" in dealing with both indented and plain material, for he expressly said: "Some bowls [his Type I] show traces of coiling."

Spier (1917: 207), referring to "coils," in dealing with corrugated ware, implied "*coiling*" for that category of pottery alone.

Morris (1917: 24–25):

> But to lay spiral coil upon coil in uniform thickness, to regulate the length of the coils so as to produce the desired slopes of the vessel walls, and at the same time to make each pressure of the thumb mold its minute portion of the design which is worked out in the coiling, demands an initial accuracy of movement, and a control of a complicated technique worthy of an artist and not at all compatible with the bungling skill of a beginner.

This equates coiling with corrugation for decorative purposes.[150]

(*supra*) and that of the Pueblo of San Ildefonso, as described and illustrated by Guthe (1925: 31 ff., pls. 12*b*, 13–15).

[144] This is perhaps true of megascopic observation; however, microscopic examination often reveals dependable criteria of technique. A medium-power, wide-field microscope should suffice for such a purpose.

[145] Linné (1925: 79) inferentially excluded the Choroti from the users of his "continuous coiling" method.

[146] Such cases perhaps may be said to illustrate tectodecorative, single or multiple fillet coiling, provided that the building process is continuous and carried on spirally.

[147] Italics are those of the present writer.

[148] Holmes (1889: 249) denied coiling and implied annular construction for the Potomac–Tidewater Virginia area.

[149] *Vide infra*, Nelson (1916); Morris (1917); Kidder (1925); and *cf.* Shepard (1936: 553, note 1).

[150] As Kidder (1936: 300, note 3) points out, it was Morris (1917) who "first called attention to the difficulty of making corrugated pottery. His paper conclusively refuted the then widely held theory that corrugated wares were older and more primitive than smooth-surfaced pottery." The significance of these observations inheres in their potential value toward chronological reconstructions in areas in which plain and corrugated coiling are in evidence. Kidder (1936: 297 ff.) is cautious and abstains

Kidder and Guernsey (1919: 141–142):

> Corrugated pottery was built up by winding round and round on itself a long, thin fillet of clay which, in well made pieces, is continuous from its beginning at the bottom of the vessel to its termination at the rim; in some large jars this fillet attains a length of more than 200 feet.[151] The laying up of this coil is in itself a very difficult matter, and when it is considered that it was often also notched, indented, waved or otherwise ornamentally modified during the building process, it will be realized how far removed this technique must have been from the first attempts of a nascent art. In the making of smooth-faced pottery any irregularity may be rubbed down or filled in, any fault of outline corrected by humoring the plastic walls into shape; in corrugated ware, however, no mistake could be corrected, and from beginning to end the coil must have been laid on with a sure hand and steady eye that must have come from long practice, not only in the handling of clay, but in its mixing exactly to the proper consistency for this delicate work. We think that it is no exaggeration to say that a large, ornamentally indented, corrugated olla required more skill for its construction than any other form of hand made pottery that has ever been produced in ancient or modern times.

This quotation furnishes an excellent illustration of the delicate procedure requisite in coiling which combines corrugation.

Kidder (1925: 7–8):

> The post-Basket Makers . . . built up their vessels by adding to the growing walls successive *rings* of clay. . . . In pre-Pueblo times it became the fashion purposely to leave unsmoothed the last few *rings* at the neck of certain small cooking vessels. . . . This was the beginning of the elaborate *coiled or corrugated* technique, later so widely used. . . . The plain, broad *rings* of the pre-Pueblo ware were replaced by a continuous thin fillet of clay applied spirally; the junctions between the successive laps of the fillets were left unobliterated, not only at the neck, but over the entire vessel; and the fillet itself was also notched or pinched or otherwise indented to produce various ornamental effects. . . . Thus was made the well known *coil or corrugated* cooking-ware so characteristic of all the archaic true Pueblo ruins." [152]

Here the "rings" may signify either several fillets used in true *coiling*, which itself is given as synonymous with corrugation, or the circuit method of construction. That the modern Pueblo of San Ildefonso use the circuit variant of building is well demonstrated by Guthe's detailed report.

Guthe (1925: 31 ff.), however, calls this process "moulding," and the base foundation "mould or puki."

> Very briefly the process is as follows: The potter first forms a pancake-shaped pat of paste from six to eight inches in diameter; this she presses into the mould or puki, to form a base.[153] Then the walls of the vessel are built up by the addition of successive ropes, or rolls, of paste laid one upon another. . . . *If the roll is not long enough completely to encircle the pat, another is made and placed on it in a similar manner. When the edge has thus been completely encircled, the unused remainder of the roll is pinched off and tossed back on the mass of paste on the canvas.* . . . When the flattening [of the applied roll] has been finished, the puki has made a second complete revolution, and *the junction of the two ends of the roll* is again directly in front of the potter. . . . *When one roll has been completely welded on, flattened out, and incorporated into the vessel, another roll is formed, and is applied in exactly the same manner.* Thus the building proceeds to the height at which the shaping is begun.[154] (Guthe, 1925: 31–35.)

There is no direct mention of coiling; yet Guthe (1925: 35–36) quotes Binns (1910: 69 ff.) for the express purpose of "showing how closely modern studio-practice, undoubtedly evolved experimentally, resembles Pueblo Indian methods."[155]

---

from direct deductions. He states: "Whether this indented corrugation preceded or came after the development of coiling is as yet unknown. Presumably, however, coiling was an outgrowth of indenting, in order that the regular sequence of the indentations should not be interrupted. At all events, both techniques made apparently an almost simultaneous appearance. The next step, the corrugating of the entire surface of the vessel, including the base, was probably taken after the practice of indenting had become well established." (*Cf.* also Kidder, 1936: 386–387, quoted *infra*.) With clarification of the constructional methodology in pottery-making it should become possible to establish the relative time position of the several variants of coiling, and of the circuit mode not only in the "Anasazi ceramic family" (Kidder, 1936: 590), but elsewhere as well. In the case of the East, specifically with the Catawba, the ring building process is to be included in such considerations.

[151] Such a fillet is composed of a considerable number of individual links.

[152] Italics are those of the present writer. The implications of this quotation suggest somewhat of a nucleus for chronological differentiation of the several constructional processes involved. Indeed they seem to provide a tempting lead with which to pursue further inquiries regarding the time relationship of circuit building, combination of circuits and coiling of neck, and all-over coiling. *Cf.* Kidder (1916: 255).

[153] *Cf.* Holmes (1886: 372) for the statement that this is "not a mold in the ordinary sense."

[154] Italics are those of the present writer.

[155] While Binns (1910: 70) intended to describe true coiling, and had partially done so, he did not exercise

The resulting ambiguity is misleading, as will presently be exemplified by a concrete case in point.[156] However, to continue with the Pueblo first.

Morris (1927: 198):

> By the end of the post-Basket Maker period, all types of San Juan pottery, indigenous to the area under consideration, had passed their point of origin . . . and corrugated ware, while not actually represented, was present in the banded neck pots.

These specimens should be useful for tectonical analysis and chronological placement of the method.

Bunzel (1929: 6–7):

> As a result . . . of the complete absence of the potter's wheel in pre-Columbian America and the failure of its native adaptation subsequently . . . all pottery is still made by coiling technique[157] . . . by adding narrow rounds of clay to a base set in a mold. The technique employed at San Ildefonso . . . is typical of the whole area. . . . The walls are built by adding rounds of clay.

This is another example of qualifying the Pueblo circuit building method as "coiling."

Kidder: (1936: 297–298):

> To understand the significance of this peculiar treatment [*i. e.*, "the practice of surface embellishment by manipulation of the structural coil"—thus involving corrugation] requires discussion of the so-called "coiling" method of pottery making. . . . To achieve larger pieces and to provide them with restricted orifices it is necessary to build upward from a prepared base (either made with the hands or pressed into a mold) by the addition of successive increments of clay.[158] Although this may be done by placing lumps at intervals upon the edge,[159] a much more effective way, and one which has been and is still being used by almost all peoples in the pre-wheel stage of pottery making, is to build by means of a small roll of clay sufficiently long to encircle the whole edge and thus raise it evenly.[160] Such is the coiling method. The term is really a misnomer, for to coil means to manipulate cylindrically or spirally a continuous element. "Ring building" would more accurately describe the process as generally carried out, even in the Southwest.[161] But in a part of that area, during later Developmental Pueblo times, it became customary to construct cooking vessels of a single strand of clay, which began at the base and, coiling upon itself spirally, continued to the orifice. The strand was of course not made its full length before work began; it was added to from time to time; but so carefully was each new increment joined that it is seldom possible to perceive a junction. The coiling element may therefore properly be said to have been continuous.[162]

Kidder continues (1936: 386–387):

> The so-called "coiling" method of making pottery has been described by Holmes,[163] Morris,[164] and others;[165] and, as employed by the modern potters of San Ildefonso, by Guthe.[166] Briefly, it consists of the building of a vessel by the addition to its growing wall of successive ring-like strips of clay; or, *and this is true coiling, by laying up spirally, a single long strip.* This general method was probably used by the prehistoric Pueblos and by their cultural ancestors, the Late Basket Makers. I say probably because the junctions between the rings or between the laps of the coil were normally removed from all vessels except cooking pots by scraping and smoothing. Upon the exterior of cooking pots the coils were often allowed to remain; and were manipulated decoratively. But that most, if not all, non-culinary ware was also built

---

sufficient discrimination when he said: "A roll of clay is taken, one end laid in the center of the bat and the rest is coiled around it in a spiral line. [Thus far, this is true coiling.] . . . *In raising the walls it is best to pinch off the roll of clay when one circle has been completed and the new roll should be begun at another point so that all the joints will not be at the same place. This plan is better than coiling a long roll in a spiral for in this case one side of the piece will be higher than the other.*" (Italics are those of the present writer.) No doubt can exist as to the meaning of the last two sentences quoted: the "better plan" is none other than circuit building, expressly preferred to spirally carried coiling.

[156] *Cf.* the citations of Gifford (1928), *infra.*

[157] Bunzel (1929: 6) allows the exception of modeling only.

[158] This statement disregards sectional building achieved by means other than coiling. *Cf.* the Catawba method of ring or circuit erection of two or more portions of the body and their subsequent joining to produce a complete form.

[159] *I. e.*, construction by the addition of either amorphous dabs (Kidder's lumps) or specifically shaped morsels of paste.

[160] This is exactly the chief determinant of circuit building, *i. e.*, that mode of procedure in wall erection in which each individual roll of paste equals, in its length, the circumference of the growing vessel at the plane upon which the fillet is being applied. *Cf.* the Catawba procedure.

[161] This would then be confused with the ring-building variant of annular construction.

[162] There follows footnote 1: "Regarding the above statement Miss Shepard comments: 'I believe the two ends were welded in place on the pot. . . . Actually, I have found a good many examples of these welded ends.'"

[163] *Cf.* my quotation, *supra.*

[164] *Cf.* my quotation, *supra.*

[165] *Cf.* my quotations of Cushing, Harrison, Linné, and Wissler, *supra.*

[166] *Cf.* my quotations of Guthe, *supra.* Note Kidder's direct reference to "coiling" at San Ildefonso as against Guthe's "moulding."

up with strips is strongly indicated both by the clean horizontal fractures which sometimes occur; and, even more clearly, by the presence of unobliterated *rings* in the interiors of certain pieces with orifices too small to have permitted introduction of the hand. *All modern Pueblo pottery is also built with rings; it is therefore to be inferred that the same practice obtained for the decorated wares of Pecos.* There is, indeed, some actual evidence to that effect, and the fragments which yield it, also provide interesting information as to how the work was done.[167]

Miss Shepard (1936: 552) writes of corrugated ware:

In many types of Pueblo culinary ware the structural coils have been carefully preserved and regularly dented. . . . It is possible to recover much of the detail of construction by close observation and systematic experimentation.[168]

Miss Shepard (1936: 440, 441) uses the term "moulding" in a collective constructional sense, whereas with reference to "Moulding and Shaping" of the modern Pueblo potters she says (1936: 448): "Guthe's description of pottery making at San Ildefonso furnishes an excellent example of the observations which should be made in recording methods in building *coiled* ware."[169] Kidder, as already quoted, clearly stressed the difference between coiling and "ring" building. However, even in his "Discussion" (1936: 590) we read: "Anasazi pottery was made by coiling."[170]

In order to illustrate the process usually called "coiling with paddle and anvil," I quote from a pertinent work of Rogers (1936: 9), dealing with the technique of the Southern Diegueño:

Coils average about twelve inches in length regardless of the size of the piece under construction. . . . The [first] coil is fastened to the base by pushing a small section of it over and on to the outer wall of the base with the left thumb, at regular intervals of one inch, while the right hand feeds the coil in a clockwise direction. . . . If the coil is not sufficiently long to go around the base once it is spliced with part of another coil. . . . Spiral coiling, as in the Puebloan technique, is unknown; *and each coil is a unit concentric to its predecessor. After* the bonding process is completed the coil is beaten flat with the paddle.[171]

Obviously, the tectonic principle of this method depends on the circuit variant of annular construction, the unit being either a single or a composite fillet. The paddling itself is a distinctly separate step which follows the placement, adjustment, and bonding of the fillet. The potter, then, does not "coil with the paddle," but merely beats the fillet flat *subsequent* to the constructional manipulation. The tectonic components are not coils in the true sense of the word.

A revealing example of confusing the process of construction proper with subsequent manipulation, appears in Spier (1928: 139):[172]

Cylinders, 2 cm. in diameter, are rolled out between the palms and coiled in a clockwise direction on a tray basket, patted on top the while with a small, smooth stick to make them adhere. . . . When the flat coil reaches a diameter of 20 cm. the sides are built up by coiling in the same manner; the potter supporting them with a smooth, round pebble inside while the exterior is paddled.

This citation shows that the basal part is built by plano-coiling. The use of the stick, for the purpose of bonding, is clearly a post-constructional process. The wall construction, although specified as "coiling in the same manner," could not, obviously, be attained by plano-coiling. Spier's

[167] Italics are those of the present writer.

[168] There follows an excellent account of Miss Shepard's detailed studies of the manufacture of indented and corrugated wares of Pecos, which deal with the criteria of the original manipulation. In footnote 1, p. 553, Miss Shepard justly comments: "There is some looseness in the use of the term *coil* and *corrugated*. They are here used in the following sense: coil refers to a process in pottery making [the process obviously implying construction, *vide infra*]. A vessel is considered coiled if it is built by successive rolls of clay placed *either in rings or spirally* [emphasis added] irrespective of subsequent smoothing and without reference to the method of shaping, whether by hand, a gourd rind tool or with paddle and anvil." The stipulated segregation of construction from shaping is noteworthy. On the other hand, it is apparent that Miss Shepard was not immune to the general confusion regarding constructional methods, for she included annular manipulation under coiling.

[169] Italics are those of the present writer. It will be recalled that Guthe did not qualify the process as coiling.

[170] This, the author stipulates, distinguished the Anasazi ceramic family from the Hohokam group, which is said to be characterized by the paddle and anvil method. It seems in point to note that the distinction rests on the consideration of two wholly separate steps of manufacture which involve entirely different principles of manipulation, *i. e., construction proper* (Anasazi) and *post-constructional conditioning* (Hohokam). Coiling as a tectonic procedure is well recognized in Hohokam pottery (*cf.* Roberts, 1937: 20). Its paddle and anvil technique, now generally accepted, does not preclude the possibility of coiling or of annular building. (*Cf.* V. J. Fewkes, 1938: 197; 1941*b*.)

[171] Italics are those of the present writer.

[172] *Cf.* also V. J. Fewkes, 1941*b*: 162–193.

text does not allow the reader to perceive the true tectonic principle involved. Yet it may be *presumed* that either an annular or a true coiling method was observed by Spier.

Gifford (1928), in his study of the Southwest, conceived of two major categories of pottery technique: "coiling without the paddle and anvil"; and "coiling with the paddle and anvil."[173] His work furnishes an example of the danger of applying the term "coiling" in a loose sense. "Coiled pottery made without the paddle and anvil," says Gifford (1928: 354), "is characteristic of Pueblo culture, both ancient and modern,[174] and is also made by the Navaho."[175] The original source on the Navaho, cited by Gifford, reads as follows:

> The paste is rolled out into long, slender pieces or ropes; this done, a flat, round cake of the desired circumference is made of a lump of the mud, and serves as the bottom of the pot around which *one of the rolls of mud is wound and made fast by pressing and gently kneading with the fingers. . . . Another roll is added and fastened in the same way, by which process the potter is enabled to give the pot the desired shape and size.*[176]

This description plainly identifies the process as circuit building. Speaking of his own observations among the Cocopa, who use the paddle and anvil—*in thinning the wall*—Gifford writes (1928: 355): "The . . . cylinder of clay . . . *proved to be just of the right length for one circuit of the edge of the growing vessel. . . . The process of adding concentric coils was continued until the vessel was complete.*"[177] This in itself is ample proof of circuit building, and not coiling.[178]

Gifford (1928: 355) conditionally assumed coiling for the Paiute, deriving his source from Lowie's brief reference (1924: 225) to the subject. At the time of Lowie's investigations, the Paiute were no longer actually making pottery, although the craft was in the memory of one woman whose recollections Lowie recorded (1924: 225). Gifford's assumption appears to be based on the following passage (Lowie, 1924: 225): "The coiled technique was employed." From the vague statements given by Lowie's informant it is not possible to reconstruct the probable building method followed by the Paiute. Nor can the circumscribed reference to "coiled technique" be used, since it lacks proof of the diagnostic criteria. Lowie, however, refers (1924: 226) to specimens of Shoshonean pottery in the Peabody Museum, Harvard University, and in the Museum of the American Indian, Heye Foundation, New York.

Through the courtesy of the Peabody Museum, I have examined the vessels in Cambridge, and am able to furnish the following data regarding their history and manufacture. There are three vessels in the collection, all labeled Paiute. Two of these, each "restored from old sherds," are both numbered 12131. The third, numbered 9448, is described in the catalogue entry as a "cooking pot, to sit in sand, made in imitation of old pots." All were collected by Dr. E. Palmer, the "imitation" vessel in 1875, and the "old sherds" in 1887. There are no notes relating to this pottery in the records of the Peabody Museum aside from the brief notations accompanying the catalogue entries. The restoration of the two vessels from the "old sherds" was done in Cambridge by the late Mr. S. J. Guernsey. The restored specimens, one with approximately one-quarter, the other with approximately one-half of plaster work, as well as the intact "imitation" pot, are of a roughly conical shape, each with an obtusely pointed bottom. The two restored vessels reveal sundry positive evidence which indicates the following manner of construction. The base appears to have been modeled in plain hands and shaped within a support made of a piece of skin, the gathered lower terminal and vertical folds of which left clear impressions on the bottom. The body of the wall was built by exterior molding over a form consisting either of a weighted skin or a carrier type basket smeared with some substance (pitch, potter's paste?) to attain smoothness. The paste medium appears to have been added in the form of irregular strips, the overlaps and bonding of which are plainly in

---

[173] *Cf.* also V. J. Fewkes, 1941*b*: 163.

[174] Citing Goddard (1927: 86), apparently the following passage: "Round after round of clay, rolled into a slender cylinder, is applied." This suggests circuit building, which is certainly characteristic of the modern Pueblo technique.

[175] There follows a lengthy quotation from Guthe (1925), which has already been discussed. Gifford accepted Guthe's "moulding" as "coiling" without comment.

[176] *Franciscan Fathers* (1910: 289); italics those of the present writer.—*Cf.* also Tschopik (1941: 23 ff.) and Keur (1941, especially pp. 52 ff.).

[177] Italics those of the present writer. Gifford (*ibid.*) expressly identified the *wall-thinning* function of the paddle process. Gifford (1933: 318 ff.) reiterated this in a tantalizingly stinted account of Cocopa pottery-making. *Cf.* V. J. Fewkes (1941*b*: 163).

[178] The presence or absence of the paddle and anvil does not affect the manipulation incidental to the placement, adjustment, and bonding of the tectonic components. In handling the paste medium during these processes, the potter's hands must be free. *Cf.* Gifford (1933: 318 ff.) for an illustration of this principle.

evidence. Finger stroking on both surfaces is clearly identified by broad streaks and minute striations. The specimen having only about one quarter of plaster restoration shows a complete continuity of its wall from base to rim. Its measurements are: maximum height, 0.393 m.; maximum diameter, taken inwardly at the rim, 0.315 m.; average wall thickness, 0.006 m. The "imitation" pot apparently had its basal part and approximately the lower third of its body modeled in hands, whereas the rest of the wall was constructed by adding irregular morsels of paste, but without the use of a mold. Overlapping of the paste additions, bonding, and finger stroking, are again fully in evidence. The interior surface, however, appears to have been smoothed down with some soft vegetable fibres, perhaps a bunch of grass blades, which left well-defined traces aligned in more or less horizontal or oblique bands.[179] *There is no evidence of coiling in any of the three vessels.*

Under his distribution of "the non-paddle method of coiling," Gifford (1928: 356) included the Catawba. As has been shown, the contemporary Catawba are totally unacquainted with coiling. And there is no evidence to indicate that they ever used it in former times, or that they acquired it in the course of their contacts with the Cherokee. Harrington's account (1908*a*) of Catawba pottery-making, which Gifford cites (1928: 356), does not stipulate the circuit process, but the reader cannot fail to recognize its existence.[180]

Gifford (1928: 357) accepted Harrington's find (1922: 194, fig. 33) of a pottery anvil in Tennessee as a sufficient indication that the ancient Cherokee practiced "coiling" of the paddle and anvil variety.[181] For the modern Cherokee, Gifford (1928: 357) also followed Harrington (1908*b*). It will be recalled that Ewi Katalsta, the Cherokee potter observed at work by Harrington 1908*b*: 223), did not demonstrate coiling,[182] but circuit building.[183]

Gifford (1928: 357) also misinterpreted the interesting constructional procedure reported by Lothrop (1927), by referring to it as "an unusual combination of *shaping and coiling*,[184] with the bottom of the pot formed last. In this method no paddle or anvil is used, the entire *shaping* being done with the hands."[185] The Guatajiagua example, unique in several respects,[186] depends essentially on hand modeling. Only in the formation of the bottom and sometimes, though not always, in the adjustment of the rim, does the potter employ the circuit technique;[187] she then uses either a single or a composite fillet, which does not exceed the circumference of the vessel at the plane of its attachment.[188] There can be no doubt, however, that the principle of coiling is totally absent.

The foregoing citations should suffice for the present purpose. My next concern is to suggest a classification of manual construction methods in primitive pottery-making. This is done here in full cognizance of my personal limitations in coping with the task. The conceptions are rooted largely in empirical criteria, and there is no intent whatsoever to propose any set standards. Admittedly, my own observations of potters at work provide the most useful sources; these are augmented by dependable literary records. In dealing with the archaeological material, only obviously recognizable criteria have been taken into consideration. It must be stressed that I am here concerned primarily with the phenomena and tasks of the North American continent.

[179] This vessel, while plastic, cracked in drying; the defect was somewhat alleviated by an interesting method of mending which consisted of applying potter's paste, of the same nature as that used in the original construction, over and along the rift.

[180] Harrington stated (1924: 271): "I was interested then [1908] to observe that the Catawba used the coil method." This was based on Harrington, 1908. Some of the potters at Catawba still remember Harrington's field work at the reservation, and they agree that coiling was not practiced within their memory. This is significant, for Mrs. Sally Gordon, now past seventy years of age, corroborated the statement; her integrity and dependability are well established.

[181] The wooden paddles illustrated by Harrington (1922: fig. 33) are modern specimens. Harrington did not express his opinion as to the manufacturing technique in either of his two groups of archaeological pottery.

[182] Which, it is true, is characteristic of the aboriginal Cherokee method, but which no longer survives with the contemporary Cherokee potters.

[183] Griffin (1935: 19–20) considered the Catawba and Cherokee methods as "identical examples of coiling."

[184] Not specified by Lothrop.

[185] Italics those of the present writer.

[186] *E. g.*, the constructional procedure, the difficult posture maintained by the potter, the guidance of shaping achieved with the aid of her toes, and the final step taken in closing the small orifice of the bottom (*cf.* Lothrop, 1927).

[187] Lothrop (1927: 113) expressly states that *the rim* is "built up by adding coils." This, apparently, induced Gifford (1928: 357) to speak of the technique of the Guatajiagua potters as a "combination of shaping and coiling"—actually modeling supplemented with circuit and morsel building.

[188] Personal information from Dr. Lothrop.

Empirically, spatially, and temporally viewed, there are two broad possible manners of constructing a pottery vessel *exclusively by the hands:*

A. By modeling, wherein the potter produces the vessel by direct shaping from a lump of paste, without necessarily using any additional material (fig. 3).
B. By segmental building, wherein the potter uses one or more tectonic components, in the form either of a fillet or of a morsel, depends on a progressive and orderly procedure, and uses additional paste as the desired vessel grows (fig. 5).

The principle involved in modeling is self-evident from the stated definition. Two main categories of product may be attained by modeling:

I. Vessels, *i. e.*, receptacles capable of accommodating certain contents.
II. Appendages, such as lugs, handles, spouts, decorative features (*e. g.*, knobs, nodules, warts, ribs, ridges), and figurines (anthropomorphic, theriomorphic, etc.). These may have:
    1. A solid body.
    2. A hollow body.

The methods of segmental construction may be classified as follows:

I. Methods depending on fillet components.
    1. Annular procedure.
        (*a*) Circuit variant.
            (i) Single fillet (fig. 5).
                Examples: modern Catawba practice (*vide supra*); modern Pueblo practice (Guthe, 1925: 31 ff.).
            (ii) Composite fillet.
                Example: modern Yuma practice (Rogers, 1936: 9).
        (*b*) Ring variant (fig. 6).
            Examples: modern Catawba practice (*vide supra*); modern Maya practice (Thompson, 1934: 95).
    2. Coiling.
        (*a*) Single(?) fillet.
            A possible although questionable example, the alleged method of the eighteenth-century Natchez (Butel-Dumont, 1753, **2**: 271 ff.).[189]
        (*b*) Linked (chained) fillet (fig. 32).
            Example: Anasazi corrugated ware (Kidder-Shepard, 1936: 297–298, 552–553).
II. Methods depending on morsel components.
    Examples: Much of the ware made by the modern descendants of the Gay Head Indians, Martha's Vineyard, Massachusetts (V. J. Fewkes, 1941*a*: 67). Perhaps also some of the archaeological laminated pottery of the Woodland, *i. e.*, that made without the use of a mold (Fewkes, 1937: 145).

The scheme just outlined (*cf.* also V. J. Fewkes, 1940: 142) is, of course, subject to modification and elaboration. Its framework, however, has proved itself adequate in sampling a series of literary sources for classificatory purposes. In some of these, details concerning the diagnostic characteristics of constructional technique were found to be either recondite or ambiguous. Yet with very few and rather insignificant exceptions it has been possible to interpret these cases in terms of the classification just discussed.

Coiling in pottery-making is not a very simple process. Whether its origin can be traced to basketry technique[190] or not,[191] the manipulation does not suggest a rudimentary stage in pottery manufacture. It is, of course, imperative to agree with Morris (1917: 24 ff.) and with Kidder and Guernsey (1919: 141 ff.) that corrugation represents a signal advance in coiling. Logically considered, the plain, hand-modeling mode of construction appears to be the most elementary; empirically, it is perhaps to be viewed as the most simple, most widely spread, and most profusely surviving method of pottery-making. And it is the modeling technique which best serves initial efforts in pottery-making—again empirically viewed. The circuit building, it seems to me, is also a rather simple procedure; I should have no hesitation in regarding it as an antecedent of coiling in the Southwest if the sample material which I have examined in this relation is representative enough. Ring building, on the other hand, impresses the observer as a specialized development.

It would be futile to attempt to rationalize the likely historical priority of modeling over circuit building, for it cannot be done on an equal temporal and spatial basis.[192] Such a question is

[189] Positive proof of a historical example of the indubitable use of a single fillet does not seem to exist.

[190] *Cf.* Mason (1902: 160; 1911: 98).

[191] *Cf.* the caution expressed by Morris (1917: 29).

[192] Certain regions, however, such as the Southwest, do offer a more promising ground in this respect.

inseparable from that regarding the origin of pottery itself. That opinions are nowise in accord with respect to this perennially interesting but virtually intangible subject, is all too well known. The technologist, it appears, rather inclines to support multiple origins, sometimes perhaps quite to the discomfiture of the diffusionist. The consternation of the opposing side —that favoring a single origin—is not substantially alleviated by recourse to convergence, the alleged criteria of which are often superficial and irreconcilable with spatial and temporal aspects. In any event, it seems that it is not so much the technologist who insists on his view, but often rather his interpreter.

The student of pottery technology is concerned not with matters which escape tangible approach, but with positive criteria and objective methodology. Pottery studies are, by the very nature of the subject, endeavors of a composite and diverse character; they inevitably necessitate a host of heterogeneous considerations. It is imperative to be as technical as necessary, for in dealing with the complex material the student must be exacting and exhaustive. The signal value of pottery in culture history has too often been stressed to require a review in the present writing. It was not until very recently that technological research has really been realized as a distinct aid to archaeology. But a full dose of the new discipline, still in an experimental stage, is not a palatable one for many an archaeologist; some are baffled thereby, others quite distrustful as to how far the details should be pursued. Whatever the reaction, the new discipline seems to be here to stay, for its utility—although yet to be tested in several respects—is, on the whole, quite patent. It is particularly desirable that the future work be directed along the lines of a broad approach rather than concentrated on disproportionate details. In other words, less of a vertical and much more of a horizontal expansion in technological research should prove more beneficial to archaeology.

The constructional aspect is a constituent part of the technology of pottery. As such it deserves attention commensurate with that devoted to other attributes of a given pot. Moreover, it lends itself to an objective determination; more precisely so, of course, upon direct observation of the process, but also, with growing success, in the study of fragments.

It is in response to such thoughts as have just been touched upon that I have gone into the matter of coiling at this length. In as much as the Catawba do not, and to my best knowledge never did, resort to coiling, it was imperative to include a critical examination of the subject.

BIBLIOGRAPHY

Ashley, M. E. 1932. A study of the ceramic art of the Etowans. *Etowah Papers* **3**: 107 ff. New Haven.

Barber, E. A. 1878. Anthropological news. *Am. Antiquarian* **1** (2): 110 ff.

Binns, C. F. 1910. The potter's craft. New York.

Bunzel, R. L. 1929. The Pueblo potter. New York.

Bushnell, D. I., Jr. 1906. The Sloane collection in the British Museum. *Am. Anthropologist* **8** (4): 671 ff.

—— 1909. The Choctaw of Bayou Lacomb. *Bur. Am. Ethnol. Bull.* 48.

—— 1934. Tribal migrations east of the Mississippi. *Smithsonian Misc. Coll.* **89** (12).

Butel-Dumont de Montigny, G. M. 1753. *Memoires historiques sur la Louisiane* [etc.]. 2 v. Paris.

Claflin, W. H., Jr. 1931. The Stallings Island Mound, Columbia County, Georgia. *Papers Peabody Mus. Am. Archaeology and Ethnology* **14** (1).

Cushing, F. H. 1886. A study of the Pueblo pottery as illustrative of the Zuni culture growth. *Bur. Am. Ethnol., Fourth Ann. Rept.*: 467 ff.

—— 1894. The germ of Shoreland pottery. *Mem. Internat. Cong. Anthropology:* 217 ff. Chicago.

Davidson, D. S. 1935. Notes on Slaughter Creek. *Bull. Archaeol. Soc. Delaware* **2** (2): 1 ff.

Du Pratz, Le P. 1758. *Histoire de la Louisiane* [etc.]. 3 v. Paris.

Farabee, W. C. 1922. Indian tribes of eastern Peru. *Papers Peabody Mus. Am. Archaeology and Ethnology* **10**.

Fewkes, J. W. 1909. Ancient Zuni pottery. Putnam Anniversary Volume: 43 ff. New York.

Fewkes, V. J. 1937. Aboriginal potsherds from Red River, Manitoba. *Am. Antiquity* **3** (2): 143 ff.

—— 1938. Review of "Pottery of Pecos, II, by A. V. Kidder and A. O. Shepard, New Haven, 1936." *Am. Jour. Archaeology* **42** (1): 195 ff.

—— 1940. Methods of pottery manufacture. *Am. Antiquity* **6** (2): 172 ff.

—— 1941*a*. A preliminary report on pottery making at Gay Head, Martha's Vineyard, Massachusetts. *Am. Antiquity* **7** (1): 67 ff.

—— 1941*b*. The function of paddle and anvil in pottery making. *Am. Antiquity* **7** (2), pt. 1: 162 ff.

Ford, J. A. 1936. Analysis of Indian village sites collections from Louisiana and Mississippi. *Louisiana Dept. Conservation, Anthropol. Study* 2.

Ford, J. A., and G. R. Willey. 1941. An interpretation of the prehistory of the Eastern United States. *Am. Anthropologist* **43** (3): 325 ff.

Franciscan Fathers. 1910. An ethnologic dictionary of the Navaho language. Saint Michaels, Arizona.

Gifford, E. W. 1928. Pottery making in the Southwest. *Univ. California Publ. Am. Archaeology and Ethnology* **23** (8): 353 ff.

—— 1933. The Cocopa. *Univ. California Publ. Am. Archaeology and Ethnology* **31** (5): 257 ff.

—— 1935. Pottery making in the Southwest. *Am. Anthropologist* **37** (3): 535.

GODDARD, P. E. 1927. Indians of the Southwest. Ed. 3. *Am. Mus. Nat. Hist. Handbook Ser.* 2.

GREGORIE, A. K. 1925. Notes on Sewee Indians and Indian remains of Christ Church parish, Charleston County, South Carolina. *Contr. Charleston Mus.* 5.

GRIFFIN, J. B. 1935. Aboriginal methods of pottery manufacture in the Eastern United States. *Pennsylvania Archaeologist* **5** (1): 19 ff.

GUERNSEY, S. J. See Kidder, A. V., and S. J. Guernsey.

GUTHE, C. E. 1925. Pueblo pottery making, a study at the village of San Ildefonso. *Papers Phillips Acad. Southwestern Exped.* **2**. New Haven.

HARRINGTON, M. R. 1908*a*. Catawba potters and their work. *Am. Anthropologist* **10** (3): 399 ff.

—— 1908*b*. The last of the Iroquois potters. *New York State Mus. Bull.* 137: 222 ff.

—— 1922. Cherokee and earlier remains on upper Tennessee River. *Indian Notes and Monographs, Misc. Ser.*, 24. New York.

—— 1924. An ancient village site of the Shinnecock. *Anthropol. Papers Am. Mus. Nat. Hist.* **17** (5): 227 ff.

HARRISON, H. S. 1928. Pots and pans—the history of ceramics. *The Beginning of Things* **8**. London.

HOLMES, W. H. 1886*a*. Pottery of the ancient Pueblo. *Bur. Am. Ethnol., Fourth Ann. Rept.*: 257 ff.

—— 1886*b*. Ancient pottery of the Mississippi Valley. *Bur. Am. Ethnol., Fourth Ann. Rept.*: 361 ff.

—— 1889. Pottery of the Potomac tide-water region. *Am. Anthropologist* **2** (2): 246 ff.

—— 1903. Aboriginal pottery of the Eastern United States. *Bur. Am. Ethnol., Twentieth Ann. Rept.*

KEUR, D. L. 1941. Big Bead Mesa [etc.]. *Mem. Soc. Am. Archaeology*, suppl. to *Am. Antiquity* **7** (2), pt. 2.

KIDDER, A. V. 1915. Pottery of the Pajarito Plateau and of some adjacent regions in New Mexico. *Mem. Am. Anthropol. Assoc.* **2** (6): 407 ff.

—— 1916. The pottery of the Casas Grandes district, Chihuahua. Holmes Anniversary Vol.: 253 ff. Washington.

—— 1925. "Introduction" to C. E. Guthe (1925), *q.v.*

—— 1936. Pottery of Pecos, II. (Includes A. O. Shepard: Technology of Pecos pottery.) *Papers Phillips Acad. Southwestern Exped.* **7**. New Haven.

KIDDER, A. V., AND S. J. GUERNSEY. 1919. Archaeological explorations in northeastern Arizona. *Bur. Am. Ethnol. Bull.* 65.

KIDDER, A. V., AND A. O. SHEPARD. 1936. See A. V. Kidder, 1936.

LEDERER, J. 1672. The discoveries of John Lederer, in three several marches from Virginia to the west of Carolina [etc.]. London.

LINNÉ, S. 1925. The technique of South American ceramics. *Göteborgs Kungl. Vetenskaps- och Vitterhets-Samhälles Handlingar Fjärde Följden* **29** (5).

LOTHROP, S. K. 1927. The potters of Guatajiagua, Salvador. *Indian Notes* **4** (2): 109 ff.

LOWIE, R. H. 1924. Notes on Shoshonean ethnography. *Anthropol. Papers Am. Mus. Nat. Hist.* **20**: 185 ff.

MASON, O. T. 1877. Anthropological news. *Am. Naturalist* **11**: 624 ff.

—— 1902. The origins of inventions. A study of industries among primitive peoples. New York.

—— 1911. Woman's share in primitive culture. New York.

MOONEY, J. 1894. The Siouan tribes of the East. *Bur. Am. Ethnol. Bull.* 22.

—— 1900. Myths of the Cherokee. *Bur. Am. Ethnol., Nineteenth Ann. Rept.*

MORRIS, E. H. 1917. The place of corrugated ware in southwestern pottery. *Am. Anthropologist* **19** (1): 24 ff.

—— 1927. The beginnings of pottery making in the San Juan Area. *Anthropol. Papers Am. Mus. Nat. Hist.* **28** (2): 125 ff.

MYER, W. E. 1928. Two prehistoric villages in Tennessee. *Bur. Am. Ethnol., Forty-first Ann. Rept.*: 485 ff.

NELSON, N. C. 1916. Chronology of the Tano ruins, New Mexico. *Am. Anthropologist* **18** (2): 159 ff.

PENNYPACKER, S. W., 2nd. 1937*a*. Friendship as a functional motive in ceramic types of eastern North America. In: Twenty-fifth Anniversary Studies: Philadelphia Anthropological Society: 147 ff.

—— 1937*b*. A note on Catawba ceramics. *Pennsylvania Archaeologist* **7** (3): 55 ff.

POLLARD, J. G. 1894. The Pamunkey Indians of Virginia. *Bur. Am. Ethnol. Bull.* 17.

RIDGWAY, R. 1912. Color standards and color nomenclature. Washington, D. C.

ROBERTS, F. H. H., JR. 1935. A survey of Southwestern archaeology. *Am. Anthropologist* **37** (1): 1 ff.

—— 1936. A survey of Southwestern archaeology. *Ann. Rept. Smithsonian Inst.* 1935: 507 ff.

—— 1937. Archaeology in the Southwest. *Am. Antiquity* **3** (1): 3 ff.

ROGERS, M. J. 1936. Yuman pottery making. *San Diego Mus. Papers* **2**.

SCHURIG, M. 1930. Die Südseetöpferei. Leipzig.

SHEPARD, A. O. 1936. See Kidder, A. V., Pottery of Pecos, II, 1936.

SMYTH, J. F. D. 1784. A tour in the United States of America. London.

SPECK, F. G. 1907. Some outlines of aboriginal culture in Southeastern United States. *Am. Anthropologist* **9** (2): 287 ff.

—— 1909. Ethnology of the Yuchi Indians. *Anthropol. Publ. Univ. Mus., Univ. Pennsylvania* **1** (4).

—— 1913. Some Catawba texts and folklore. *Jour. Am. Folklore* **16**: 319 ff.

—— 1914. The Nanticoke community of Delaware. *Contr. Mus. Am. Indian, Heye Foundation*, **2** (4).

—— 1925. The Rappahannock Indians of Virginia. *Indian Notes and Monographs* **5** (3): 25 ff.

—— 1928. Chapters on the ethnology of the Powhatan tribes of Virginia. *Indian Notes and Monographs* **1** (5): 225 ff.

—— 1934. Catawba texts. *Columbia Univ. Contr. Anthropology* **24**.

—— 1935. Siouan tribes of the Carolinas as known from Catawba, Tutelo, and documentary sources. *Am. Anthropologist* **37** (2): 201 ff.

—— 1939. Catawba religious beliefs, mortuary customs, and dances. *Primitive Man* **12** (2): 21 ff.

SPIER, L. 1917. An outline for a chronology of Zuni ruins. *Anthropol. Papers Am. Mus. Nat. Hist.* **18** (3): 207 ff.

—— 1928. Havasupai ethnography. *Anthropol. Papers Am. Mus. Nat. Hist.* **29** (3): 81 ff.

STEVENSON, J. 1883. Illustrated catalogue of the collections obtained from the Indians of New Mexico and Arizona in 1879. *Bur. Am. Ethnol., Second Ann. Rept.*: 307 ff.

STEVENSON, M. C. 1904. The Zuni Indians. *Bur. Am. Ethnol., Twenty-third Ann. Rept.*

SWANTON, J. R. 1911. Indian tribes of the lower Mississippi Valley and adjacent coast of the Gulf of Mexico. *Bur. Am. Ethnol. Bull.* 43.

—— 1923. New light on the early history of the Siouan peoples. *Jour. Washington Acad. Sci.* **13** (3): 33 ff.

—— 1931. Source material for the social and ceremonial life of the Choctaw Indians. *Bur. Am. Ethnol. Bull.* 103.

—— 1935. Notes on the cultural provinces of the Southeast. *Am. Anthropologist* **37** (3): 373 ff.

—— 1936. Early history of the eastern Siouan tribes. In: Essays in anthropology presented to A. L. Kroeber: 371 ff. Berkeley.

THOMPSON, J. E. 1930. Ethnology of the Mayas of southern and central British Honduras. *Field Mus. Nat. Hist. Publ.* 274, *Anthropol. Ser.* **17** (2).

TIMBERLAKE, H. 1765. The memoirs of Lieut. Henry Timberlake (who accompanied three Cherokee Indians to London in the year 1762). London.

WISSLER, C. 1938. The American Indian, ed. 3. New York.

RECOVERY OF MIDWESTERN PRAIRIES FROM DROUGHT

JOHN E. WEAVER

University of Nebraska

(*Read November 19, 1943*)

THE long period of desiccation, which began in 1934, finally terminated in 1940. The western portion of True Prairie was greatly decimated and a veritable mosaic of types of vegetation resulted. These ranged from well-preserved relict communities of bluestems (*Andropogon*) to mixed stands of grasses on nearly bared soil, but included communities of western wheat grass (*Agropyron smithii*), short grasses (*Bouteloua gracilis* and *Buchloe dactyloides*), needle grass (*Stipa spartea*), and prairie dropseed (*Sporobolus heterolepis*). These types varied in size from a few square rods to many acres (fig. 1). One 15-acre prairie, which was mapped in detail, revealed five types, each of which was dominated by a single species, and eight mixed types where two species in each were in definite control. Some of these types of vegetation, moreover, recurred in several parts of this prairie.

FIG. 1. A well-drained level lowland near Lincoln, Nebraska, formerly completely covered with big bluestem. Relatively small amounts of this type still remain (foreground); western wheat grass invaded the drought-bared soil and formed distinct, pure communities (right center); and the needle grass type (light) developed in the background.

Over the entire western margin of True Prairie, wide areas of soil were often still open to invaders in 1940 and others were populated only thinly by individuals of a potential grassland type. Even the vegetation in relict communities had usually been greatly thinned by the drought, and dynamic development involved in mutual invasions, competition, and reaction was pronounced (fig. 2). In this process of revegetation western wheat grass played an important part.

FIG. 2. Former little bluestem prairie which was greatly damaged by drought but naturally reseeded to needle grass, which is now the climax type.

Invasion of western wheat grass into bared soil populated only by early (weedy) stages of the subsere was well advanced during the drought. When good rains came, it continued rapidly along the periphery of its established communities as well as into neighboring ones not yet stabilized. This marginal advance into new territory with a larger supply of water was demarked by greater stature, increased production of foliage, and more numerous flower stalks than in soil already occupied. Under competition for water with wheat grass, relict individuals of nearly all species of grasses and forbs were greatly dwarfed; they often failed to flower, and ultimately succumbed. Hence, wheat grass communities formed remarkably pure stands (fig. 3). Blue grama (*Bouteloua gracilis*) and buffalo grass (*Buchloe dactyloides*), and under certain

FIG. 3. Portion of a bluestem prairie at Carleton, Nebraska, where the former grasses have been almost completely replaced by western wheat grass.

rainfall distribution side-oats grama (*Bouteloua curtipendula*), alone could successfully adjust themselves to this competition.

Blue grama and its seedlings were so much more drought-resistant than all other species that this grass spread rapidly from relatively small predrought communities even during years of low rainfall. It greatly increased the area of these communities and produced many new ones. After 1940 vegetation was rapidly thickened by the production of new bunches between the old ones and often by the spreading of the stolons of buffalo grass as it recovered under the increased rainfall. The stolons advanced several feet in a single season. They threaded their way among the bunches of blue grama and thus formed a dense sod. Buffalo grass also spread into bared soil and formed a sod. Western wheat grass very successfully invaded communities of blue grama. The rhizomes at first advanced in the bare soil between the bunches or sodded mats but later spread thinly through them. Not only were buffalo grass communities similarly invaded but invasions were also mutual, since seedlings of both short grasses regularly established themselves and later developed an understory to the taller-growing wheat grass. Such a layering of grasses at two levels occurs typically only in mixed prairie (fig. 4). This phenomenon of the formation of mixed prairie by mutual invasion of short grasses and mid grass has occurred over hundreds of square miles. The eastward shift of one major plant community or association into the territory of another association is one of the most striking results of the drought.

Needle grass, characteristic of dry habitats, showed during the dry years a continuous spreading into bared or semibared upland and even onto low ground. The deep planting of the seed and, like wheat grass, early growth in spring were helpful in evading drought. Needle grass formed an important predrought grassland type which has increased in area at least tenfold. The dwarfing effect of needle grass upon neighboring species, while less than that of wheat grass, is still pronounced. The process of thickening its stand, with increase in water supply since the end of the drought, has proceeded so rapidly and the older bunches have become so well established that the climax community has been produced and may remain as such for many generations.

The prairie dropseed, formerly of such small importance that its type held sixth place in predrought True Prairie, has benefited greatly by the death of its competitors and consequent opening of new territory for invasion. Cessation of mowing the prairie, because of very little forage production during the driest years, permitted this late-blooming bunch grass to ripen abundant seed, and its territory has consequently been enormously increased. It has often nearly stabilized its communities almost to the exclusion of other species. Elsewhere more scattered individuals constitute an important component of the mixed type of vegetation.

These several communities occupy most of the

FIG. 4. A thin stand of western wheat grass which has recently invaded an area already occupied by buffalo grass and blue grama and has thus formed mixed prairie.

grassland and stand out clearly like pieces of a mosaic—western wheat grass, blue grama or buffalo grass, mixed prairie, needle grass, prairie dropseed, and relict bluestems. But over considerable areas, widely scattered plants of various species recovered, reseeded, or invaded at about the same time. This gave rise to mixed communities of variable composition. It is here that the battle for dominance will be waged fiercely and probably for a long time. Such mixed communities of needle grass, blue grama, big bluestem (*Andropogon furcatus*), prairie dropseed, tall dropseed (*Sporobolus asper*), and other grasses more or less equally distributed have no place in stabilized True Prairie. With the reseeding of all species since the drought and with vegetative propagation proceeding rapidly, competition has become severe, reactions have become marked, and the nature and manner of stabilization have become a phenomenon of great scientific interest. For it is here that the fitness or unfitness of the species for its habitat, as determined by its reactions and coactions, will be revealed.

Finally, completing the prairie cover are those relict areas where the bluestem grasses were only thinned and not destroyed. These communities, like the preceding, vary greatly in size. They are the portions least disturbed by invasion, although in many cases over 80 percent of little bluestem (*Andropogon scoparius*) succumbed. It was slowly replaced by the rapid spread of big bluestem by rhizomes, by a remarkable increase in side-oats grama, June grass (*Koeleria cristata*), and other native species formerly a part of the community. The process was relatively slow before the end of the drought, but has been accelerated greatly since. Where little bluestem was nearly or completely exterminated, big bluestem sometimes increased from a 10 percent stand to form a pure dense cover (fig. 5).

FIG. 5. An open cover of big bluestem resulting from the death by drought of the more shallowly rooted little bluestem at Carleton, Nebraska.

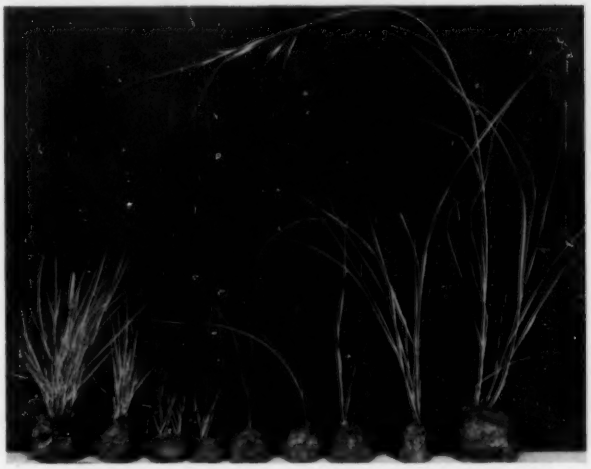


FIG. 6. Various stages in growth of June grass (left) and needle grass (right). The smallest plants are of the current year's growth (samples were taken from the field in June); the largest with flower stalks are completing their third summer of development.

In 1940 much of the prairie was open to invasion, since the cover, except wheat grass, usually presented many open spaces, a square foot to several square feet in area, unoccupied by perennial grasses. These supported a stand of weedy annual grasses or other weeds, or presented a bare surface for seedlings. Debris was not plentiful and the usual thin mulch was not present to protect the soil.

After 1940 seedlings of most species of grasses became common, and were very plentiful by 1943 (fig. 6). Many survived and gradually filled the interspaces everywhere. By tillering profusely, small bunches of grass extended their control over many times that of their predrought basal area. Individual postdrought seedlings grew into bunches several inches in basal diameter. By rhizomes and by stolons the bared spaces were being repopulated by a vigorous vegetation of unusual stature. This increase in development and size was best shown where repopulated denuded areas were adjacent to little-disturbed ones of sod-bound grasses of the same kind which had been neither removed by drought nor succeeded by younger plants.

Corresponding destruction and return of forbs add an interesting part to the process of recovery. Many forbs decreased 50 to 90 percent, others entirely disappeared, but a few increased greatly. Some returned only as seedlings from seed maintained in the soil through the six or seven years of drought. Others grew from rhizomes, root crowns, bulbs, or corms still alive in the soil. Not infrequently the first year's growth from these long-dormant parts consisted of a single small shoot easily mistaken for the stem of a seedling. Revival from dormancy sometimes did not occur until one or more years after good rains came. Some species have not yet returned. Vegetation is usually slow and conservative in its changes.

Restoration of the plant cover is far less advanced toward a complete sod in some portions of True Prairie than in others. This varies with the degree of depletion, the types of vegetation concerned, and with the amount and seasonal distribution of water in the soil. For deficient water content of soil and not temperature or humidity is the chief cause of drought. Availability of water is also the chief condition for recovery. The main patterns of recovery are similar everywhere, but the methods vary greatly with species and season. A new factor has only recently been discovered in great longevity of underground parts of certain grasses.

The uniformity and purity of stands of wheat grass during drought (in the absence of short grasses) have been noted. But only in 1943 —the third summer of good rainfall—have certain wheat grass prairies begun to change to mixed wheat grass–big bluestem by the midsummer appearance of widely spread stems and small clumps of drought relicts. The spark of life remained in parts of these half-decayed root crowns and rhizomes of big bluestem at least five to seven years. Some buds were still in deep dormancy when they were unearthed in July. Little bluestem is likewise putting forth a few culms, most of which are bearing seed, from similarly long-dormant root crowns. Only in coming seasons can one evaluate the importance of this method of recovery on the composition of the climax vegetation.

The present mosaic cover of the prairie, even where complete, is not climax. Only the period of occupation of all the territory has neared completion. Where all available soil is taken, to gain a place in the sun and to receive its portion of water, a newcomer must replace established plants which normally live for many years. This is a slow and difficult procedure. Return to the predrought dynamic equilibrium of climax types will necessitate many changes. If normal precipitation prevails over a sufficiently long period and this return does occur, little bluestem will increase to first rank as regards abundance. The needle grass type will be greatly reduced. Short grasses will almost disappear, and the prairie dropseed type will become infrequent. Wheat grass will be reduced from its present rank of the most important prairie dominant to less than 1 percent of the prairie mixture.

A knowledge of competitive abilities of the various species and how they thrive or wane in mixtures is not only of much scientific importance but also finds immediate practical application. The change from bluestems to wheat grass, for example, at once reduced the value of the grass crop for grazing and hay. Water infiltration has been greatly retarded, water loss by runoff increased, and the excellent structure of normal prairie soil has gradually deteriorated. In True Prairie, cattle graze wheat grass only when better forage is greatly depleted, and it is grazed closely only when other vegetation is absent. Yield of hay in a year of good precipitation (1943) was reduced by invasion of wheat grass in three widely separated upland prairies to 59 percent of that of big bluestem. Comparison of yields were made with big bluestem growing only a few feet distant from the wheat grass. Extensive experiments have shown that water infiltration in soil covered with wheat grass sod was only 42 percent as rapid as its entry into the same kind of soil clothed with big bluestem. Wheat grass is not a desirable species to be employed in alternate husbandry, that is, in rotation of cultivated crops every few years with grass intermixed with legumes, in order to regain good soil structure and to improve fertility by increasing nitrate content and organic matter.

It should be emphasized that the mosaic of grassland patterns at the end of the drought or even after their adjustment during three years of normal rainfall represents only a phase in recovery. Many years with good precipitation will be required to build into the grassland its layered structure, its societies of composites, legumes, and other forbs, and for the less xeric grasses to attain the great abundance and wide distribution characteristic of these dominants under the usual climatic control.

The drought has shown clearly that nature has

richly endowed True Prairie with many species, some of which are best adapted to cover the soil, enrich it, and hold it against the forces of erosion during moist climatic cycles. Others which are then found in such small amounts that they seem almost a non-essential part of grassland rapidly increase to great abundance and become of great importance when a severe drought cycle occurs. This is what happened in the 1934–1940 drought and must have occurred many times in the historical and geological past, although no written record has been made.

After a six-year survey (1928 to 1933) of upland True Prairie of eastern Nebraska and adjacent states, 50 percent of the area was assigned to little bluestem alone. Conversely, western wheat grass was scarcely ever abundant, and side-oats grama and annual fescue (*Festuca octoflora*) occurred sparingly. But when the terrible drought struck suddenly in 1934, little bluestem was nearly swept away, big bluestem and others greatly diminished in abundance, and the soil was nearly bared of vegetation. Nature at once sowed the land with annual fescue, and certain forbs became widely spread weeds in the temporary cover. As if by magic wheat grass covered the land and side-oats grama soon attained second place among prairie grasses. In the dry hot soil, needle grass, prairie dropseed, and other xeric grasses spread widely. In grazed prairie, short grasses migrated eastward and a wave of sand dropseed (*Sporobolus cryptandrus*) swept over the drought-bared land. Thus, it seems that the wide variations in grassland climate are well within the range of endurance of prairie species. Such great changes and adjustments as have been made are among the species of prairie only. Vegetation from neither desert nor forest has been permitted to invade.

In mixed prairie of western Kansas and eastern Colorado, native grassland still occupies about one-half of the area. Practically all of it is used for pasture. These vast range lands are now recovering from the terrible destruction wrought by overgrazing, drought, and dust-burial during the past decade. Grass is man's most powerful ally in soil and water conservation, and pastures are tremendously important as the cheapest and most satisfactory feed for cattle. Hence, the recovery of the western range is not only a phenomenon of outstanding ecological interest but also one of major economic importance. This study is a joint investigation by Dr. F. W. Albertson, of Ft. Hays Kansas State College, and the writer.

The blue grama - buffalo grass type is of greatest extent and importance in western Kansas and eastern Colorado. Hence it received major consideration in a study of the effects of drought, dust, and intensity of grazing on cover and yield of short-grass ranges.

An understanding of present conditions can be had only by a knowledge of past treatment of the ranges. High wind movement was common throughout the drought period and especially severe during the spring of 1935. Soil which had been pulverized by excessive cultivation was carried as dust in large quantities from cultivated fields and deposited in layers or drifts on the drought-stricken prairie plants. The cover of native vegetation was thus greatly reduced, in fact, the grasses were almost or completely annihilated over vast areas of range land. Planted crops failed to grow because of deficient moisture. This compelled the rancher to increase the usual length of the grazing season as well as the intensity of range utilization. This extreme overgrazing was responsible for further decline in the amount of original plant cover. The soil when devoid of native vegetation was often populated by numerous weedy annuals.

The prairie soil was bared by adverse environmental conditions and became extremely susceptible to wind erosion. It was the source of dust for further great storms which greatly damaged the ranges. When pastures were placed under the stress of these unfavorable environmental conditions, it became evident that in those where good management had been practiced the original plant population suffered far less than where overgrazing or overgrazing with intense dusting (a local term meaning covered with dust) had occurred.

At four widely separated stations, Hays, Ness City, Dighton, and Quinter, Kansas, pastures were selected for study at the end of the most severe drought ever recorded. Four classes of pastures were investigated at each station where the range use was known or where records were obtained. Class 1 had been lightly grazed and lightly dusted; class 2 heavily grazed and lightly dusted; class 3 lightly grazed but heavily dusted; and class 4 both heavily grazed and heavily dusted (fig. 7).

Increase in basal area of perennial grasses, amount of the various species concerned, and dry weight of forage produced were the chief

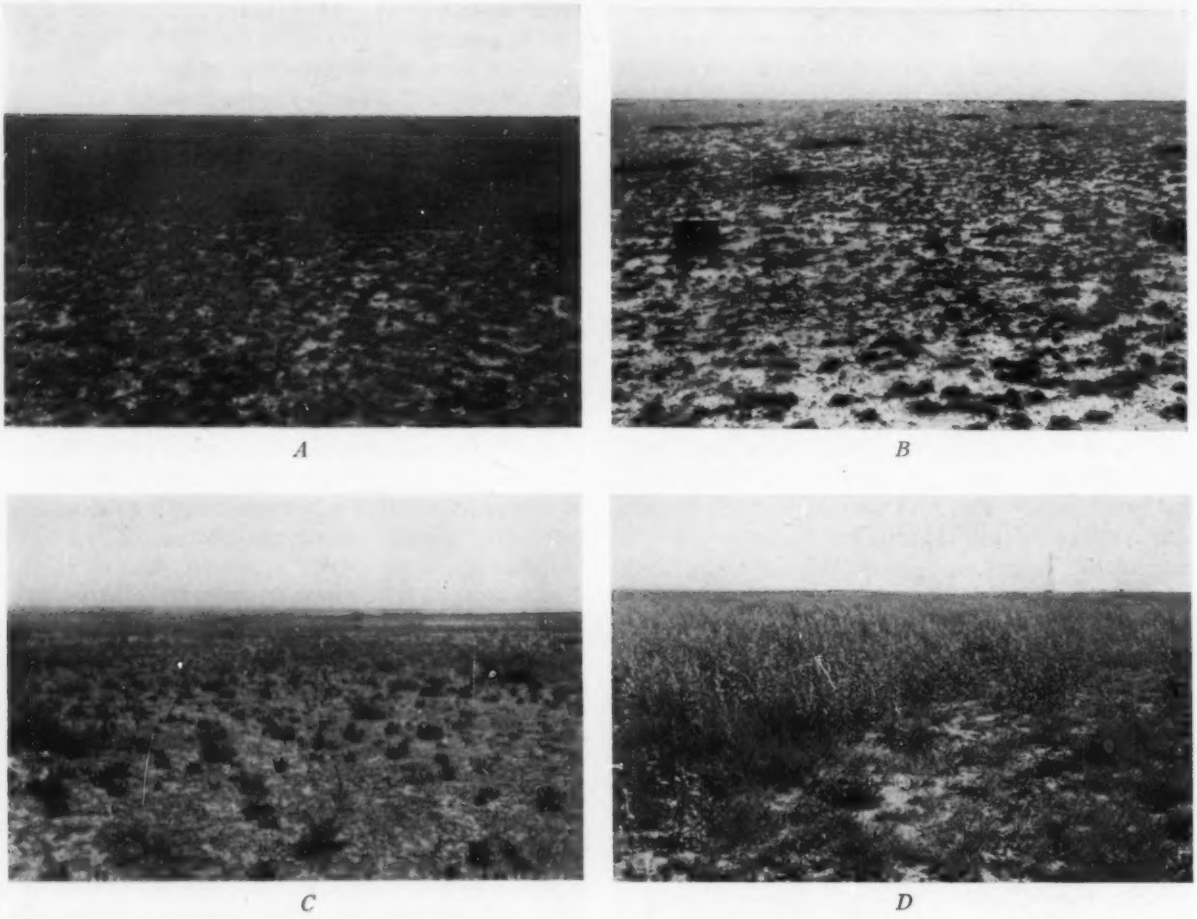


FIG. 7. *A*. Typical view of a class 1 pasture in midwestern Kansas where the open cover remained intact throughout most of the drought. *B*. A class 2 pasture with typical open cover. In early spring such ranges are usually infested with annual weeds. *C*. Class 3 pasture where the original amount of cover has been much reduced and its composition greatly modified. *D*. An overgrazed and heavily dusted (class 4) pasture. Most of the original vegetation died and had been replaced by lamb's quarters, rough pigweed, and sunflower. All photos taken in September, 1940.

criteria used in measuring recovery. Exclosures against stock were made. In each exclosure 20 permanent meter quadrats were clipped annually, and the basal area of the perennial grasses was charted by means of a pantograph in 1940 and 1941. Forbs were so few as to be almost negligible, but annual weeds were abundant. Increase or decrease of basal area of each species of perennial grass was ascertained, and the separate yields of short grasses and weeds were obtained.

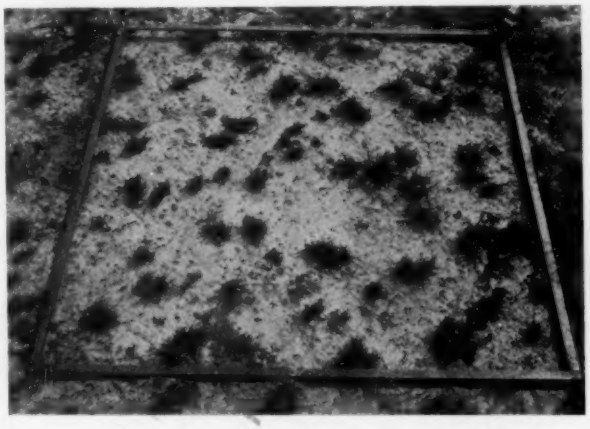
Low precipitation in 1939, 7 to 10 inches below the normal 23 to 25 inches, and an extremely dry fall had further reduced the much depleted plant cover. Drought also weakened the vegetation so greatly that it remained almost dormant at all stations the following spring, and nearly throughout the summer at Dighton and Quinter.

The growing season of 1940 showed much improvement over that of 1939, but precipitation was below normal at all stations during April and June. At Hays and Ness City only was the amount of precipitation normal. Despite the moderate rainfall, the thoroughly dried soil was often without available water at all stations and growth was not continuous and yields were not high, except those of annual weeds.

In 1941 the rainfall was above normal more than half of the time from April to September. The soil was wet from early spring until July; periods without rain were neither prolonged nor severe, and the drought was definitely broken.

Basal area was first ascertained in the fall of 1940 after a year of good recovery, since only then was it possible to separate with certainty

dormant plants from those that were dead. It was invariably least in all class 4 pastures at any station, and nearly always highest in class 1. Cover in pasture classes 2 and 3 was intermediate, but often higher in the lightly grazed but heavily dusted class 3. In fact, a light cover of dust often furnished some protection against drought by aiding dormancy. The average basal cover of short grasses in the four class 1 pastures was 18 percent; that in the others in order was 13.7, 13.4, and 2.2 percent. These percentages are strikingly low compared with the 80 to 95 percent basal cover which was usual before the drought. During the excellent season for growth in 1941 these pastures extended their basal cover to 56.8, 46.3, 44.2, and 17.2 percent, respectively.



A

B

FIG. 8. *A*. View of short grass in class 1 pasture at Dighton, Kansas, in the fall of 1940. The basal cover is about 11 percent. *B*. The same quadrat as it appeared early in July, 1942. Buffalo grass had increased greatly and the basal cover was 80 percent.

These results illustrate the wonderful recuperative powers of both blue grama, the most drought-resistant species of the midwest, and buffalo grass, which spreads vegetatively with great rapidity (fig. 8). Although seedling grasses played a very minor role in the process of recovery, yet pasture classes 1 to 3 on an average tripled their cover and basal cover in class 4 increased nearly eightfold.

Blue grama at first exceeded buffalo grass in amount of basal area in all but two class 1 and 2 pastures. But at the end of 1941, buffalo grass exceeded blue grama in seven of these eight pastures. There was about 4.5 times as much buffalo grass as blue grama in the average of class 1 pastures, and about 3 times as much in the average of class 2. In the heavily dusted pastures (classes 3 and 4) in 1940, buffalo grass ranked somewhat higher in all except two, but percentage of cover was low. In 1941, so great were its gains, it ranked far ahead of blue grama, often being 3 to 5 times as abundant. Despite fourfold increases of cover in low-grade pastures, several good years for seed production and growth of seedlings will be required for complete restoration of the range. In many pastures both spread and yield of the perennial grasses were greatly reduced by an abundance of weeds, chiefly little barley (*Hordeum pusillum*), peppergrass (*Lepidium densiflorum*), lamb's quarters (*Chenopodium album*), and Russian thistle (*Salsola pestifer*). They absorbed much water, and greatly reduced the light both when living and when dead. They were most abundant in class 4 pastures. Weeds usually yielded one-fourth ton to more than one ton per acre.

Only broad correlations were found between amount of basal cover and yield. The drought-stricken vegetation varied greatly in vigor; the amount of dust accumulation resulted in varying degrees of dormancy and protection from drought. Moreover, occurrence and amount of soil moisture available to growth were extremely variable.

Previous experience with drought and previous grazing treatments profoundly affected the vigor and future behavior of short grasses, as did also a period of protection. The latter may have been furnished by partial burial by dust. It has been noted that a year of great vegetative activity of buffalo grass, as expressed in abundant stolon production, may be followed by one with much less vegetative increase. Nor is abundant tiller production compatible with the greatest yield of seed of blue grama.

SOME VIRUSES TRANSMITTED BY AGALLIAN LEAFHOPPERS

L. M. BLACK

The Rockefeller Institute for Medical Research, Princeton, New Jersey

(*Read April 23, 1943*)

INTRODUCTION

In a recent paper (Black, 1941*b*) data were presented indicating that of the two closely related leafhoppers, *Aceratagallia sanguinolenta* (Provancher) and *Agallia constricta* Van Duzee (fig. 1), only the former transmits the New York variety of the potato yellow-dwarf virus, *Marmor vastans* Holmes var. *vulgare* Black (Holmes, 1939; Black, 1940), and only the latter transmits the New Jersey variety, *Marmor vastans* Holmes var. *agalliae* Black (Black, 1941*b*).[1] Each of these leafhoppers is the most abundant and widespread species in the genus to which it belongs. It seemed logical to suspect that other agallian leafhoppers might be carrying other potato yellow-dwarf viruses, as yet undiscovered, and that these might be disclosed by feeding insects in genera closely related to *Aceratagallia* and *Agallia* on suitable hosts. To test this hypothesis, the leafhopper *Agalliopsis novella* (Say) (fig. 1) was selected for special study because this species is the dominant one in the only other North American genus of agallian leafhoppers which is distinct from *Aceratagallia* and *Agallia*.

Fig. 1. Agallian leafhoppers. Upper row, females; lower row, males. Left to right in each row: *Aceratagallia sanguinolenta* (Provancher), *Agallia constricta* Van Duzee, and *Agalliopsis novella* (Say). × 5.6 (Photograph by J. A. Carlile.)

As a result of collecting *Agalliopsis novella* in the field and testing it on crimson clover (*Trifolium incarnatum* L.), two apparently new viruses were discovered. Attempts were made to ascertain the relationship of the two new viruses to the two potato yellow-dwarf viruses previously known. This was deemed of importance because of the possibility that the relationships of the insect vectors were paralleled by the relationships of the viruses which they transmit. So far it has not been possible to prove close relationship of the four viruses, but close relationship has been indicated by study of vectors and symptoms.

EXPERIMENTS

DISCOVERY OF VIRUSES TRANSMITTED BY *Agalliopsis novella*

Being unfamiliar with the characteristics and habitat of *Agalliopsis novella* except through literature, the writer enlisted the collaboration of Dr. P. W. Oman, who kindly consented to instruct him in the collection and recognition of this species. Accordingly, Dr. Oman and the writer spent October 30 and 31, 1941, collecting nymphs of *A. novella* in the vicinity of Washington, D. C. The weather was unfavorable and the two days of collecting resulted in the capture of only 47 nymphs. These were distributed on three crimson clover plants growing in 4-inch pots covered with celluloid cages and were transported to Princeton, N. J. The insects were transferred to a second set of three test plants on November 8, and to a third set on December 11. The survivors were removed from the third set of test plants on January 22, 1942. Dr. Oman identified the insects as be-

[1] For convenience, the New York potato yellow-dwarf virus will henceforth be referred to in this paper as the NY virus and the New Jersey potato yellow-dwarf virus as the NJ virus.

longing to the type of *A. novella* illustrated by figure 14*G* in his 1933 paper (Oman, 1933). The crimson clover plants were observed for symptoms, but none was detected. However, some of the nymphs obtained from the second set of test plants produced a new disease in 6 out of 90 crimson clover plants. Later it was discovered that other insects from the stock culture derived from the Washington collection were able to produce the malady. The disease could easily have passed unnoticed in the plants on which the original 47 insects were tested, because the plants were grown in the winter when the symptoms are milder and because they had to be discarded early owing to fungus infection. It therefore seems probable that this new virus was present in the *A. novella* nymphs collected in the vicinity of Washington, D. C., and that this insect is carrying the virus in nature in that area.

Because the outstanding symptom of this disease in clover is the enlargement of the veins, it is proposed that the disease be called "clover big-vein" and the virus "clover big-vein virus." For convenience, the virus will be referred to in this paper as the BV virus.

In an attempt to verify the natural occurrence of the BV virus in *Agalliopsis novella*, 23 nymphs were collected between September 22 and October 8, 1942, from the grounds of the Rockefeller Institute, Princeton, N. J. Each nymph was fed separately on a single crimson clover seedling from the day of collection until November 2. On that date each nymph was transferred to a second crimson clover seedling and fed on it until December 1. After removal of the insects, the test plants were observed in a greenhouse. Six of the first lot of test plants and two of the second became diseased, and the distribution of the diseased plants showed that 8 of the 23 nymphs, or 35 percent, were viruliferous when captured. None of 38 corresponding control plants became diseased. But the disease was not that caused by the BV virus. Instead, it was a distinctly new disease, and the experiment indicated that the causal virus is being carried in nature by *A. novella* in the vicinity of Princeton, N. J.

An unusual feature of this disease in clover is the delayed opening of the leaves so that many young leaves appear like little clubs. For this reason it is suggested that the disease be known as "clover club-leaf" and the virus as "clover club-leaf virus." For convenience, the virus will be referred to in this paper as the CL virus.

The presence of NJ virus in *Agallia constricta* was discovered accidentally when this insect was fed on clover plants to ascertain if it transmitted the NY virus. So far as the writer is aware, the BV virus and the CL virus are the first two plant viruses to be discovered as the result of a deliberate attempt to find new viruses in their insect vectors. The success attained suggests possible use of the method in virus surveys. Plant viruses have heretofore been observed causing disease in plants, and the insect vector has had to be discovered by subsequent investigation. This is often a long and difficult task. The method by which the BV and CL viruses were found has the inherent advantage of providing the means of transmission simultaneously with the discovery of the virus. On the other hand, it has an inherent disadvantage illustrated by the fact that the plants which the BV and CL viruses are presumed to affect in nature are as yet unknown.

SYMPTOMATOLOGY

The outstanding symptoms produced by each of the four viruses in crimson clover will now be described so that the reader may visualize the similarities and differences between them. Healthy crimson clover leaves and leaves showing symptoms of the four diseases are illustrated in figure 2.

The most conspicuous symptoms caused by the NY virus consist of vein-clearing and yellowing of the younger leaves. During the summer the yellowing is usually partially replaced by an interveinal reddish-brown color on both leaf surfaces extending from the margins inwards. All parts of the plant are dwarfed. If the plant does not die in the acute stage of infection, it passes into a chronic stage in which the symptoms are very mild. There is no rusty-brown necrosis along the veins, no vein enlargement, and no cupping of the leaves.

The symptoms caused by the NJ virus are extremely variable. Only the constant features of the disease will be described here. Symptoms usually develop first on some of the older leaves. A rusty-brown necrosis along the veins is always present in some degree, and, when affected leaves die, markings along the veins still show clearly. The interveinal areas sometimes turn yellow or even a light cream color. Necrotic streaks may appear on the petioles. Affected plants are dwarfed and sometimes die in the acute stage of infection. Usually, however, they pass into a

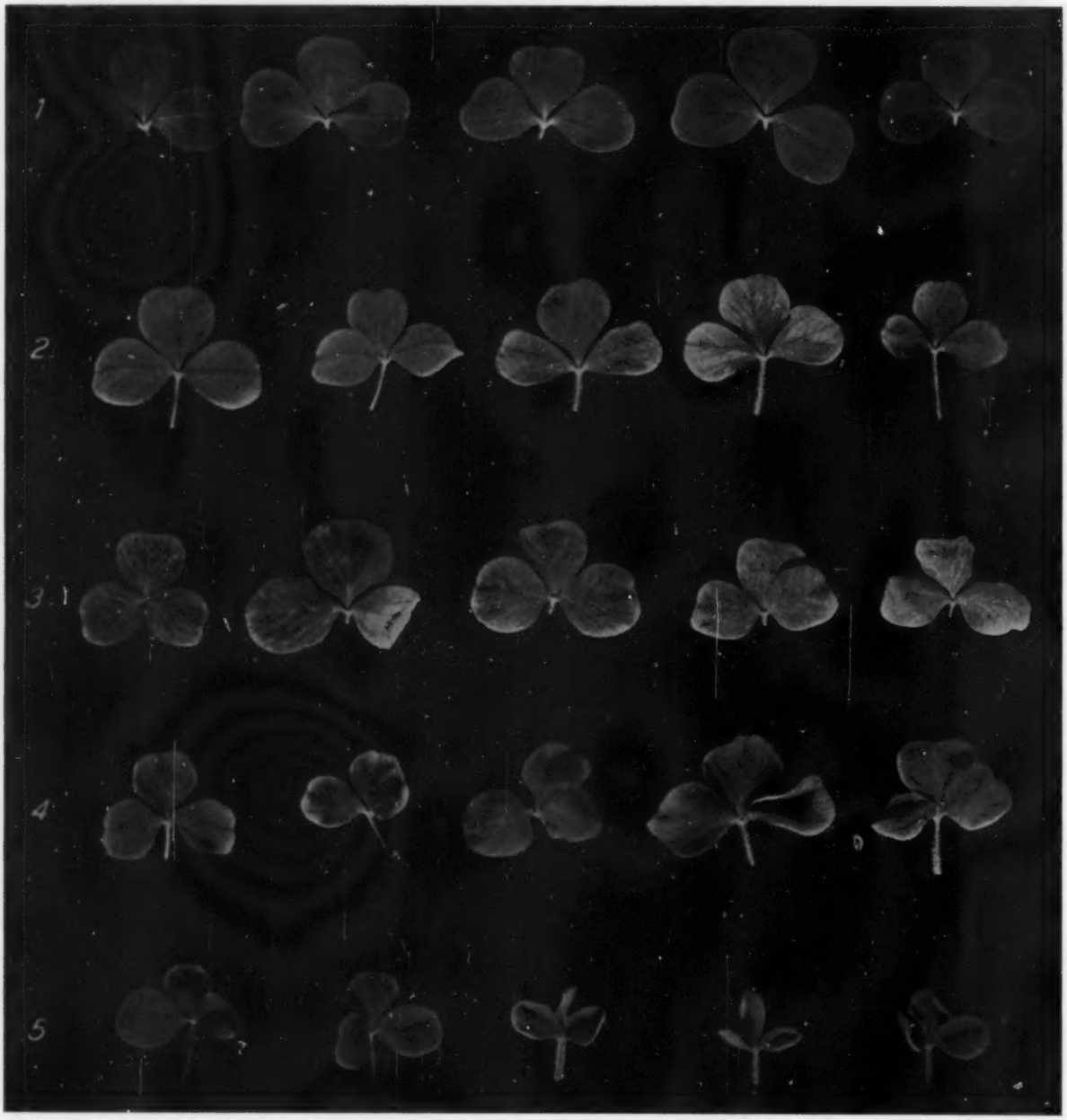


FIG. 2. The crimson clover leaves in row 1 are healthy. The different rows, 2 to 5, illustrate leaves that are infected by NY, NJ, BV, and CL viruses, respectively. Actual size (Photograph by J. A. Carlile).

chronic stage of disease in which the symptoms are mild. In this disease there is no conspicuous vein-clearing, no noticeable vein enlargement, and no cupping of the leaves.

Crimson clover plants invaded by the BV virus develop unevenly thickened veins which are depressed below the upper surface of the leaf. The enlarged veins are most readily detected by examination of the leaves from the lower surface, and they are so distinctive that positive diagnosis can be made even if only one or two veins exhibit the abnormality. Enations sometimes arise from the lower surface of these veins, and the leaves frequently curl upwards and inwards around the margin. In the summer the leaves turn yellow from the margin inwards, and later

the yellow color is partially replaced by red or purple-red. Petioles on diseased plants are undulating rather than straight. All parts of the plant are dwarfed, and the internodes are shortened so that a dense rosette results. There is no vein-clearing and no rusty-brown necrosis. The upward curling of the leaves which is frequently present is easily distinguishable from the cupping of the leaves which results from infection with the CL virus.

In the early stages of infection by the CL virus, the youngest leaves are lighter green than normal and their unfolding is delayed so that they have a clublike appearance. Some of the young opened leaves may show yellowing and dwarfing of one or more leaflets. The yellowing may appear on part of a leaflet, and in such a case it extends from the margin inwards and appears on the side of the leaflet before it appears at the tip. During the summer red or purple-red colors eventually replace the yellow around the leaf margins. Affected leaves appear narrower than normal and may be smooth or savoyed. If smooth, the leaf frequently resembles one infected by the NY virus. The infection inhibits apical growth and stimulates slightly the production of new shoots from the axils of the leaves, the new growth consisting of spindly stems and small leaves. There is no rusty-brown necrosis of the veins, no obvious vein enlargement, no obvious vein-clearing as a primary symptom.

TRANSMISSION OF VIRUSES BY *Aceratagallia sanguinolenta*, *Agallia constricta*, AND *Agalliopsis novella*

Experiments were carried out to determine the vector relationships of the viruses and the agallian leafhoppers already mentioned. Early experiments dealt with *Aceratagallia sanguinolenta* and *Agallia constricta* and the NY and NJ viruses. Later experiments included *Agalliopsis novella* and the BV and CL viruses as these became available. The general procedure in all experiments was the same. Only experiment 10 will be described in detail. Infected crimson clover plants were selected to provide sources of the different viruses. The plants were selected on the basis of symptom expression, stage of infection, and adequacy as a food source for the insects during the feeding period. The plants in 6-inch pots were covered with celluloid cages pressed down into the soil. Nonviruliferous nymphs in early instars were selected for vigor, and a mixture of the three species was added to each plant according to the following scheme:

Plants infected with	Number of plants	Insects added to each plant
NY virus	4	15s 13c 13n
NJ virus	3	20s 17c 17n
BV virus	3	20s 17c 17n
CL virus	3	20s 17c 17n

s = *Aceratagallia sanguinolenta*
c = *Agallia constricta*
n = *Agalliopsis novella*

After feeding on the diseased plants for 17 days, the insects were removed, separated according to species, and each nymph was placed on a different healthy crimson clover seedling growing in a 2½-inch pot and covered by a small celluloid cage. Each insect was tested for two weeks on each of three successive test plants unless it died in the interim. After removal of insects, test plants were placed in a greenhouse and held under observation for six weeks. The procedures in the other experiments varied only in minor details. In every experiment care was taken to see that each insect had an equal chance to pick up and to transmit the virus under test, but it was impossible to set up an exact pattern of procedure and to follow it in every particular in each experiment. Thus, the length of time insects were allowed to feed on diseased plants was the longest time that the plants provided a good source of nourishment and that the insects remained as nymphs. This period varied from 8 to 26 days in the ten different experiments and averaged 17 days. With minor exceptions, the insects were tested on three successive lots of seedlings and were allowed to feed for two weeks on each lot.

The results of these experiments are summarized briefly in table 1, which shows the proportion of insects that transmitted, and in table 2, which shows the proportion of plants infected.[2] The data of experiment 1 have already been published (Black, 1941*b*).

From tables 1 and 2 it is clear that NY virus was transmitted by *Aceratagallia sanguinolenta* but not by *Agallia constricta*. The single case of transmission by *Agalliopsis novella* may indicate rare transmission of the virus by this species. It should be pointed out that transmission of the NY virus by *A. sanguinolenta* is often more

[2] See appendix.

TABLE 1

SPECIFIC TRANSMISSION OF VIRUSES BY AGALLIAN LEAFHOPPERS AS INDICATED BY THE PROPORTION OF INSECTS TRANSMITTING. THE DENOMINATOR INDICATES THE NUMBER OF INSECTS TESTED; THE NUMERATOR IS THE NUMBER WHICH TRANSMITTED THE VIRUS INVOLVED.

Experiment	Source plant infected with	*Aceratagallia sanguinolenta*	*Agallia constricta*	*Agalliopsis novella*
1	No virus	0/33	0/33	
	NY virus	11/31	0/31	
	NJ virus	0/35	19/35	
2	No virus	0/40	0/40	
	NY virus	8/40	0/40	
	NJ virus	0/40	33/40	
3	No virus	0/32	0/32	
	NY virus	17/32	0/32	
	NJ virus	0/33	21/33	
4	NY virus	4/10		0/20
	NJ virus		12/17	0/35
	BV virus	0/49	15/35	4/36
5	NY virus		0/50	0/47
	NJ virus		17/47	0/49
	BV virus		14/47	14/46
6	NY virus	9/50		
	NJ virus	0/50		
	BV virus	0/50		
7	NY virus	2/30	0/30	0/30
	NJ virus	0/30	24/30	0/30
	BV virus	0/30	18/30	13/30
	CL virus	0/30	0/30	10/30
8	NY virus	8/30	0/30	1/27
	NJ virus	0/30	14/23	1/30
	BV virus	0/30	12/29	8/30
	CL virus	0/30	0/30	6/22
9	NY virus	9/43	0/49	0/48
	NJ virus	0/45	32/46	0/47
	BV virus	0/45	25/50	14/45
	CL virus	0/41	0/48	5/40
10	NY virus	11/50	0/42	0/48
	NJ virus	0/50	33/50	0/50
	BV virus	0/50	17/42	19/43
	CL virus	0/48	0/42	2/38
All experiments	No virus	0/105 = 0%	0/105 = 0%	
	NY virus	79/316 = 25%	0/304 = 0%	1/220 = 0%
	NJ virus	0/313 = 0%	205/321 = 64%	1/241 = 0%
	BV virus	0/254 = 0%	101/233 = 43%	72/230 = 31%
	CL virus	0/149 = 0%	0/150 = 0%	23/130 = 18%

TABLE 2

SPECIFIC TRANSMISSION OF VIRUSES BY AGALLIAN LEAFHOPPERS AS INDICATED BY THE PROPORTION OF PLANTS INFECTED. THE DENOMINATOR INDICATES THE NUMBER OF TEST PLANTS; THE NUMERATOR IS THE NUMBER THAT BECAME DISEASED BY THE VIRUS IN QUESTION.

Experiment	Source plant infected with	*Aceratagallia sanguinolenta*	*Agallia constricta*	*Agalliopsis novella*
1	No virus	0/90	0/94	
	NY virus	17/82	0/86	
	NJ virus	0/103	43/103	
2	No virus	0/90	0/95	
	NY virus	11/76	0/92	
	NJ virus	0/84	74/102	
3	No virus	0/89	0/90	
	NY virus	30/77	0/92	
	NJ virus	0/94	40/99	
4	NY virus	8/25		0/35
	NJ virus		25/38	0/71
	BV virus	0/123	27/73	4/62
5	NY virus		0/139	0/136
	NJ virus		31/131	0/140
	BV virus		22/132	17/127
6	NY virus	18/144		
	NJ virus	0/146		
	BV virus	0/144		
7	NY virus	2/89	0/89	0/87
	NJ virus	0/89	45/86	0/86
	BV virus	0/90	30/87	14/85
	CL virus	0/90	0/88	13/83
8	NY virus	12/68	0/72	1/57
	NJ virus	0/68	22/53	1/78
	BV virus	0/67	27/70	11/63
	CL virus	0/58	0/70	6/41
9	NY virus	14/123	0/131	0/133
	NJ virus	0/131	61/112	0/125
	BV virus	0/129	35/130	16/124
	CL virus	0/114	0/119	7/103
10	NY virus	19/135	0/98	0/124
	NJ virus	0/128	77/119	0/123
	BV virus	0/118	28/97	24/100
	CL virus	0/116	0/96	2/96
All experiments	No virus	0/269 = 0%	0/279 = 0%	
	NY virus	131/819 = 16%	0/799 = 0%	1/572 = 0%
	NJ virus	0/843 = 0%	418/843 = 50%	1/623 = 0%
	BV virus	0/671 = 0%	169/589 = 29%	86/561 = 15%
	CL virus	0/378 = 0%	0/373 = 0%	28/323 = 9%

efficient than in these experiments. The reason for the inefficiency here is unknown.

The NJ virus was transmitted by *Agallia constricta* but not by *Aceratagallia sanguinolenta*. These results with large numbers of insects suggest that the differential specificity of *A. sanguinolenta* and *A. constricta* towards the NY and NJ viruses may be absolute. However, the results to be described later in experiments with *Agallia quadripunctata* make one wary of concluding that this is the case. The one transmission of NJ virus by *A. novella* may indicate very occasional transmission by this insect.[3]

The BV virus was transmitted not only by *Agalliopsis novella* but also by *Agallia constricta*. Indeed, *A. constricta* proved to be a more efficient vector than *A. novella*. This is strange, because many collections of *A. constricta* at Princeton and one collection sent from Washington, D. C., by Dr. Oman have yielded only the NJ virus. *Aceratagallia sanguinolenta* did not transmit the BV virus.

[3] The New Jersey and New York yellow-dwarf diseases are very similar in potatoes. The NJ disease has been identified by insect transmission experiments in old stocks of potatoes growing on Cape May, N. J. If the virus had been discovered in such potato plants before its discovery in *Agallia constricta*, undoubtedly it would have been assumed that *Aceratagallia sanguinolenta* was the vector. The latter insect is abundant in the area. This experience suggests caution in assuming that apparently identical diseases in different geographical regions are disseminated by the same insect vector.

So far, the CL virus has been transmitted only by *Agalliopsis novella*. The fact that 8 out of 23 *A. novella* nymphs collected at Princeton, N. J., were carrying the CL virus suggests that the efficiency of transmission in this combination may be greater than indicated in tables 1 and 2. The conditions under which experiments 8, 9, and 10 were conducted may not have been optimum for this particular virus and insect.[4]

TRANSMISSION STUDIES WITH OTHER INSECTS

Fairly extensive tests of sucking insects not closely related to agallian leafhoppers have indicated that such insects do not transmit the NY virus (Black, 1937). It seemed desirable to determine whether or not other species of the genus *Aceratagallia* could act as vectors. Through the kindness of Dr. H. H. P. Severin and Dr. Norman W. Frazier, several species of this genus occurring in California were shipped to the writer. In preliminary experiments some of these were tested for transmission of the NY virus. Individual insects that transmitted the virus were later identified by Dr. P. W. Oman. As a result of this work, evidence was obtained that *Aceratagallia lyrata* (Baker), *A. obscura* Oman, and *A. curvata* Oman are capable of transmitting the NY virus. It seems probable that many species in the genus have this capacity.

Extensive experiments were carried out with the species *Agallia quadripunctata* (Provancher) and the four viruses. Young nymphs, after feeding on diseased plants, were placed in colonies of five on young crimson clover plants having two or three trifoliate leaves. In experiment 1 each colony was tested on four successive plants; in experiments 2 and 3, on three successive plants. The colony was left on each test plant for three weeks. The results are summarized in table 3. *A. quadripunctata* transmitted NY virus, NJ virus, and BV virus. It failed to transmit CL virus in these tests. A comparison of the results in table 3 with those in tables 1 and 2 shows that *A. quadripunctata*, like *A. constricta*, transmitted NJ virus and BV virus and, like *A. constricta*, it transmitted the NJ virus more efficiently than it transmitted the BV virus. However, it did not transmit the NJ virus as efficiently as did *A. constricta* and, unlike *A. constricta*, it transmitted the NY virus, although very rarely. However, transmission by two colonies in experiment 1 (table 3) occurred under circumstances which leave little doubt that this species does occasionally transmit the NY virus. All colonies, especially the two in question, were carefully examined for possible contaminating individuals of the species *Aceratagallia sanguinolenta*. Only individuals of *A. quadripunctata* were found. One of the transmitting colonies infected two out of four plants on which it was tested; the other infected three out of four. These results showed clearly that the *A. quadripunctata* colonies were doing the transmitting and that the five infected plants could not have been due to possible viruliferous *A. sanguinolenta* contaminants in the greenhouse, because, if such had been the case, the contaminations would have been scattered among the 262 test plants of the experiment. It seems probable that, out of 465 *A. quadripunctata* tested in the three experiments, only 2 transmitted the NY virus. The fact that an insect species is able to transmit a virus in such rare cases makes one cautious about concluding that *A. sanguinolenta* is never able to transmit NJ virus and *A. constricta* never able to transmit NY virus. This result also strengthens the interpretation placed on the single infection with NY virus and the single infection with NJ virus by *Agalliopsis novella* shown in tables 1 and 2. Both of these infections, it should be noted, occurred in the correct group of plants, an event that should occur once in every 16 trials by chance alone.

Another point of interest in connection with

TABLE 3

TRANSMISSION OF VIRUSES BY *Agallia quadripunctata*

Experiment	Days fed on diseased plants	Source plant infected with	Colonies infective / Colonies tested	Calculated probable % of infective insects	Plants infected / Test plants
1	15	NY virus	2/33	1	5/131
		NJ virus	15/33	11	22/131
2	23	No virus	0/30	0	0/90
		NY virus	0/30	0	0/90
		NJ virus	10/30	8	10/90
		BV virus	3/30	2	3/90
		CL virus	0/30	0	0/90
3	32	No virus	0/30	0	0/89
		NY virus	0/30	0	0/89
		NJ virus	23/30	25	32/89
		BV virus	19/30	18	29/89
		CL virus	0/30	0	0/90

[4] There is a fifth virus reported to be transmitted by an agallian leafhopper—the virus causing curly top of beets in Argentina. This virus is apparently transmitted by *Agalliana ensigera* Oman (Fawcett, 1927; Oman, 1934). Because of its transmission by an agallian leafhopper and because the symptoms it causes resemble those produced by the BV virus, it would be worth investigating for its possible relationship to the four viruses dealt with above.

the rare transmission of NY virus by *Agallia quadripunctata* is the fact that, when such transmission occurred, the transmitting insect infected more than one plant. This may be due to multiplication of the virus in such insects.

RELATIONSHIPS OF THE VIRUSES

There can be little doubt that the NY and NJ viruses are very closely related. The fact that both are readily transmitted mechanically on *Nicotiana rustica* L. and produce yellow primary lesions is highly indicative of relationship, because no other leafhopper-borne viruses are known to be transmitted as readily mechanically on any host or to produce primary lesions as a result of mechanical inoculation. The primary lesions produced by the two viruses are very similar (figs. 3 and 4), in spite of the fact that those produced by the NJ virus are usually fewer in number and develop more slowly than those of the NY virus. The systemic symptoms produced by both viruses on *N. rustica* (fig. 5), on *Solanum tuberosum* L. var. Green Mountain (fig. 6), on *Chrysanthemum leucanthemum* L. var. *pinnatifidum* Lecoq and Lamotte, and on *Trifolium incarnatum* (fig. 2) are very similar. Only minor differences can be detected in the symptoms on the first three hosts. On crimson clover the symptoms of the two viruses are similar but almost always readily distinguishable one from the other.

Moreover, three protection tests have been carried out on *Nicotiana rustica* which also provide evidence of the close relationship of these

FIG. 3. Primary lesions of New York potato yellow-dwarf virus on leaf of *Nicotiana rustica* L. 29 days after inoculation. (Photograph by J. A. Carlile.)

FIG. 4. Primary lesions of New Jersey potato yellow-dwarf virus on leaf of *Nicotiana rustica* L. 29 days after inoculation. (Photograph by J. A. Carlile.)

FIG. 5. *Nicotiana rustica* L. leaves. Left to right: healthy; systemically infected with New York potato yellow-dwarf virus; and systemically infected with New Jersey potato yellow-dwarf virus. × 0.2 (Photograph by J. A. Carlile.)

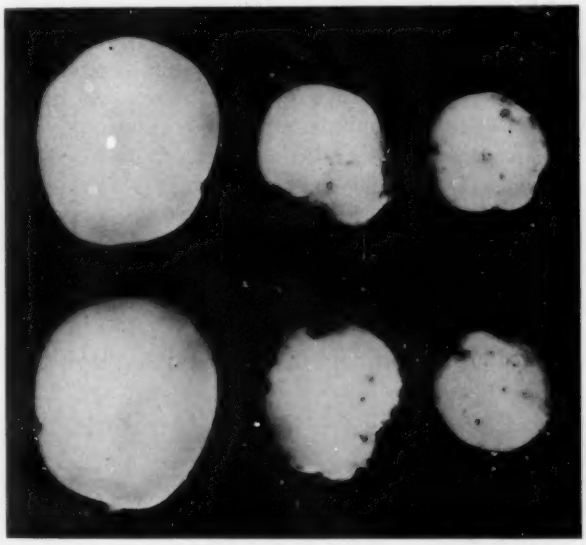


FIG. 6. Slices from potato tubers, *Solanum tuberosum* L. var. Green Mountain. Left to right in each row: healthy; infected with New York potato yellow-dwarf virus; and infected with New Jersey potato yellow-dwarf virus. × 0.7 (Photograph by J. A. Carlile.)

two viruses. In these tests a necrotizing strain (var. *lethale* Black (Black, 1940)) of the NY virus was used to inoculate *N. rustica* plants systemically invaded by NJ virus. As controls, comparable inoculations were carried out on healthy plants, on plants systemically invaded by a mild strain (strain B_2 (Black, 1940)) of the NY virus, and on plants systemically invaded by unrelated viruses. Although two of the three tests have already been briefly reported (Black, 1941*b*), all three will now be described in detail.

Six isolates of the NJ virus were obtained at different times by mechanical inoculation of *Nicotiana rustica* with juice from six crimson clover plants selected for their exhibition of various types of infection by the NJ virus. In each of the three protection tests, four *N. rustica* plants formed the experimental group. In each experiment, each of some four of these six NJ isolates was employed to inoculate four young *N. rustica* plants. As controls in each experiment, four *N. rustica* plants were held uninoculated, four were inoculated with strain B_2 of the NY virus, four with tobacco-ringspot virus (*Annulus tabaci* Holmes var. *virginiensis* Holmes), and four with either potato-calico virus (*Marmor medicaginis* Holmes var. *solani* Black and Price) (Black and Price, 1940), tobacco streak-disease virus (*A. orae* Holmes), or potato-ringspot virus (*M. dubium* Holmes var. *annulus* Holmes). The plants were pruned back and fed with an ammonium nitrate solution at appropriate intervals during the entire period of each experiment, in order to keep one shoot on a plant in a vigorous vegetative stage of growth and to prevent the plants from flowering. After good systemic symptoms of disease had developed in all the inoculated plants, all the tops were removed above the leaves judged best for inoculation. Two plants in each group of four were inoculated with the *lethale* strain of the NY virus and two were left uninoculated as controls. In experiments 1, 2, and 3, the inoculation with the *lethale* strain was performed 56, 23, and 46 days, respectively, after the first set of inoculations had been carried out. Within 21 to 33 days (average 26 days), systemic symptoms of *lethale* infection had developed on all the plants previously uninoculated or inoculated with one or another of the viruses of tobacco ringspot, potato calico, tobacco streak, or potato ringspot. In experiments 1, 2, and 3, the plants were held for observation for 62, 55, and 76 days, respectively, after inoculation with the *lethale* strain. Nevertheless, none of the six plants containing strain B_2 of the NY virus and inoculated with the *lethale* strain developed systemic symptoms of *lethale* infection, and of the 24 plants containing a strain of the NJ virus only five developed restricted *lethale* infections late in the experiment. Three of these five all contained the same isolate of the NJ virus, the symptoms of which indicated it was a poor invader. Some of the results of experiment 3 are illustrated in figure 7.

The relationships of the BV and CL viruses are not certain. Attempts to transmit BV virus to *Nicotiana rustica* by means of insects, by means

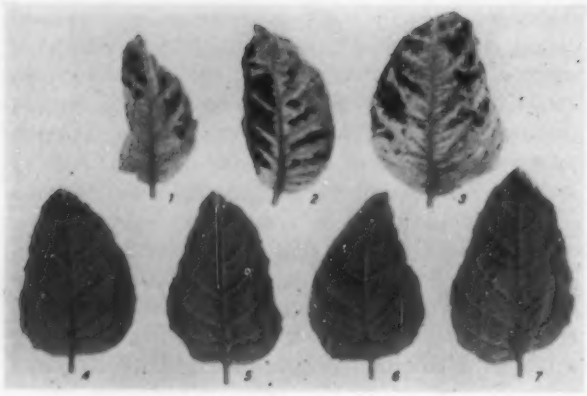


FIG. 7. *Nicotiana rustica* L. leaves representative of results obtained in protection experiments. The plants represented by leaves 1, 2, 3, 6, and 7 were inoculated with the *lethale* derivative of the New York potato yellow-dwarf virus; those represented by leaves 4 and 5 were not. When the inoculations with the *lethale* strain were carried out, the plant represented by leaf 1 was healthy; the others were systemically invaded by different viruses, namely, plant 2 by tobacco ringspot, 3 by potato ringspot, 4 and 6 by New York potato yellow-dwarf strain B_2, 5 and 7 by New Jersey potato yellow-dwarf. Leaves 1, 2, and 3 show severe systemic necrosis as a result of invasion by the *lethale* strain; the other leaves do not. × 0.3 (Photograph by J. A. Carlile.)

of dodder, and by means of rubbing leaves with juice from infected clover failed to produce any symptoms of disease in the plants. Mechanical inoculation of *N. rustica* with CL virus likewise failed. *Agallia constricta* carrying BV virus and *Agalliopsis novella* carrying CL virus failed to produce any symptoms of disease in Green Mountain potato plants on which they fed. The failure to transfer either the BV or CL virus to *N. rustica* has thus far prevented the employment of the *lethale* strain of the NY virus in protection tests. The failure to infect potato plants or *N. rustica* with the BV or CL virus may be considered as evidence against the relationship of these two viruses to the NY and NJ viruses, but it is by no means critical evidence, since host differences between virus strains are well known (Storey and McClean, 1930; Kunkel, 1932; Holmes, 1941).

Small colonies of *Agallia constricta* carrying BV virus infected 6 out of 15, and small colonies of *Agalliopsis novella* carrying CL virus infected 4 out of 12, plants grown from crimson clover cuttings chronically infected by the NY virus. Under comparable conditions, 6 out of 15 and 8 out of 12 healthy control plants became infected by BV virus and CL virus, respectively. By mechanical inoculation of *Nicotiana rustica* leaves, it was shown that neither the BV virus nor the CL virus had displaced the NY virus in the clover stem tips exhibiting symptoms of big-vein or club-leaf. However, six out of eight crimson clover cuttings chronically affected by the NY virus were infected by the NJ virus when inoculated by means of insects. In these six plants the NJ virus had an incubation period about two weeks longer than in the eight control plants, which had been healthy until each was successfully inoculated by means of insects. No such delay was observed in the case of the tests with the BV virus and the CL virus. The slight evidence of protection in this latter experiment was not deemed sufficient basis for concluding that the results with the BV virus and the CL virus meant that they were not related to the NY virus. It seems more likely that the results indicate that this type of experiment has little value in the determination of virus relationships.

A number of legume viruses, namely, pea-mosaic virus (*Marmor leguminosarum* Holmes), pea-mottle virus (*M. efficiens* Folke Johnson (Johnson, 1942)), white sweet clover mosaic virus, white clover virus 1, and white sweet clover ringspot virus, were transferred to crimson clover and their symptoms studied. The symptoms of none of these viruses resembled those of the NY, NJ, BV, and CL viruses as closely as the symptoms of the latter four viruses resembled each other.

Some of the disease symptoms and the insect vectors indicate that the four viruses are related. All four viruses cause yellowing; none causes a typical mosaic. On the other hand, the characteristic vein enlargement produced by the BV virus and the slight stimulation of axillary shoots caused by the CL virus may be interpreted as indicating lack of relationship to the NY and NJ viruses. The failure to infect *Nicotiana rustica* or potato may be similarly interpreted. In any case, the BV and CL viruses do not seem to be as closely related to the NY and NJ viruses as these two are to each other.

CLASSIFICATION AND NOMENCLATURE OF THE VIRUSES

Without attempting any final answer on the relationships and classification of the four viruses dealt with in this paper, it is the writer's opinion that for the present, because of the similarities in symptoms and because of the insect relation-

ships, they should be grouped together. Furthermore, it seems clear that, because of their vectors and their symptoms, the viruses would be better classified in the family Chlorogenaceae than in the family Marmoraceae. It is suggested that the four viruses be placed in a new genus, **AUREOGENUS** (from L. *aureus*, yellow or golden, + L. *genus*, meaning group), in the family Chlorogenaceae. It is proposed that the genus *Aureogenus* consist of viruses that have agallian leafhoppers as vectors, that cause yellowing without the mottling typical of mosaics, and that may or may not stimulate growth of axillary shoots. Acceptance of these proposals necessitates broadening the description of the family Chlorogenaceae to include symptoms caused by viruses in *Aureogenus*. It is suggested that the names for the 4 viruses should be:

Aureogenus vastans var. *vulgare* n. comb., for the NY virus

A. vastans var. *agalliae* n. comb., for the NJ virus

A. magnivena (meaning big-vein) n. sp., for the BV virus

A. clavifolium (meaning club-leaf) n. sp., for the CL virus

DISCUSSION

The specific vector relationships between the NY and NJ varieties of potato yellow-dwarf virus and the two agallian leafhoppers, *Aceratagallia sanguinolenta* and *Agallia constricta*, seem to be the first cases in which varieties of a plant virus were shown to have different specific but closely related vectors. This is clearly an example of physiological specialization of varieties of a virus with respect to closely related insect vectors comparable to the physiological specialization of races of *Puccinia graminis* Persoon var. *tritici* Eriksson and Henning with respect to varieties of *Triticum aestivum* L. It is not simply an example of one virus variety being a more virulent one than another.

Severin (1934) has reported results which may indicate that the eastern and the western varieties of aster-yellows virus (*Chlorogenus callistephi* Holmes vars. *vulgaris* Holmes and *californicus* Holmes) exhibit either physiological specialization or variation in virulence with respect to their vectors. Because of the nature of the results, one cannot distinguish between the two possibilities. *Macrosteles divisus* (Uhler) (*Cicadula sexnotata* (Fallén)) transmits both the eastern and the western varieties of the virus (Kunkel, 1932), but Severin reported that *Thamnotettix montanus* Van Duzee transmitted only the western variety. However, in the experiments in which *T. montanus* failed to transmit the eastern variety, Severin does not report the use of a successful parallel series of transmissions of the western variety as a control, and in the same paper he reports certain unexplained failures to obtain transmission of yellows by means of *M. divisus* from carrot (*Daucus carota* L.), common plantain (*Plantago major* L.), parsley (*Petroselinum hortense* Hoffm.), and parsnip (*Pastinaca sativa* L.). In the absence of controls, these failures suggest that the failure of *T. montanus* to transmit the eastern variety of the virus may not have been due to lack of ability to transmit, but may have resulted from conditions under which the experiments were carried out. At that time the inactivation of the virus in the insect (Kunkel, 1937) and in the plant (Kunkel, 1941) by high temperatures was not known. If the temperature was high at the time the transmissions were attempted, the virus may have been inactivated in the vector or in the plant. Further experiments seem desirable to test the ability of *T. montanus* to transmit the two strains of aster-yellows virus and, if Severin's report is confirmed, to ascertain whether the result is due to physiological specialization or to variation in virulence. It is possible that the western strain of the virus is more infectious than the eastern strain and can be transmitted by both *M. divisus* and *T. montanus* and that the eastern is less infectious and can be transmitted only by *M. divisus*.

It seems probable to the writer that the NY and NJ viruses are native American viruses rather than introductions from abroad. There is evidence (Barrus and Chupp, 1922) that some farmers had recognized potato yellow-dwarf for years before it was discovered by a plant pathologist in 1917. The vectors are indigenous, and the occurrence in the United States of two field varieties of the virus highly specialized in their vector relationships suggests that they evolved in this area. So far as the writer is aware, no viruses similar to the potato yellow-dwarf viruses have been reported occurring outside the Americas. If these viruses were introduced, it is necessary to suppose that each arrived as a distinct entity and established a highly specific relationship with its vector, or that after one was introduced the other evolved from it.

In the past, experiments on the specificity of insect transmission have generally tested trans-

mission by a series of unrelated insect species, usually those that feed on some economically important host of the virus. The experiments with the NY and NJ viruses reveal that the specificity of transmission may be more refined than the usual types of experiments can disclose.

It is difficult to understand the specificities of transmission in the case of the NY and NJ viruses, if the relationship between insect and virus is mechanical. Recent work (Fukushi, 1935, 1939, 1940; Kunkel, 1937; Black, 1941*a*) with certain other leafhopper-borne plant viruses indicates that they multiply in their animal vector, the insect, as well as in the plant. It is suggested that these two viruses also multiply in their vectors and that the case may be homologous with many host-parasite relationships where specificity is exceedingly refined and where the parasite alternates between hosts widely separated phylogenetically. It is worth pointing out that the NY and NJ viruses have apparently evolved along specific lines in relation to their insect vectors and not in relation to their plant hosts. Dr. S. G. Younkin informs me that thus far he has found 53 species of plants susceptible to both these viruses and none certainly immune to one of the viruses and susceptible to the other.

The fact that the two closely related NY and NJ viruses are transmitted by closely related leafhoppers suggests a parallelism between the relationships of the viruses and the relationships of their vectors. If one assumes that the insect vector of a virus is purely a matter of chance, then the NJ virus transmitted by *Agallia constricta* would have as much chance of being related to any other plant virus as to the NY virus. The fact that the NY and NJ viruses proved to be closely related seems of considerable significance. The similar type of yellowing produced by all four viruses without the mottling typical of mosaics suggests that they are related to each other. This, and the fact that all are carried by insects in the same group, is a further indication of parallelism. However, the notion of parallelism must remain no more than a hypothesis until the relationships of the BV and CL viruses are more definitely ascertained. Moreover, one cannot be certain that extrapolation of the results from one small area of this immense field will provide an accurate picture of the whole. Knowledge of the extent to which plant virus relationships are paralleled by the relationships of their vectors must depend on the accumulation of data from many investigators. The problem has an obvious bearing on questions of classification of plant viruses and on the search for insect vectors of viruses.

The possible significance of insect vectors in connection with the differentiation and classification of plant viruses has impressed several investigators. Elze (1931) proposed a classification of plant viruses based on whether or not the virus was transmitted by insects, whether or not it was specifically transmitted, and whether the incubation period was long or short. The weaknesses of Elze's proposals have been clearly presented by Kunkel (1935). Storey (1931) published an interesting speculative paper entitled "The Bearing of Insect-vectors on the Differentiation and Classification of Plant Viruses." Storey considered that most plant viruses are better characterized by their insect vectors than by the plants they affect. He pointed out the agreement between a grouping of viruses upon the type of symptom and a grouping upon the type of insect vector, mentioning the group of mosaic viruses transmitted largely by aphids, the group of yellows viruses transmitted by jassids, and a third homogeneous symptomatic group transmitted by white-flies. However, Storey wrote that the chief purpose of his paper was to point out some of the errors that might arise from the differentiation of viruses by their insect vectors, and the latter half of his paper is devoted to this end. Kunkel (1935) felt that the biological relationships of plant viruses to their vectors, when better known, might furnish a good basis for differentiation and classification. Johnson and Hoggan (1935) employed vector relationships in their descriptive key for plant viruses. Although Holmes (1939) based his classification of plant viruses primarily on the symptoms they produce in plants, four of his ten families, namely, Chlorogenaceae, Rugaceae, Savoiaceae, and Lethaceae, are also distinctively characterized by their insect relationships.

The plant viruses transmitted after an incubation period in their vectors seem to be rather precisely differentiated by their vector relationships. In fact, such differentiation seems usually to proceed to the point of identification or to the point where the virus must be one of a number of closely related strains. In the case of some viruses in this group, many different insect species have been tested without changing the picture. Additional vectors of some of these viruses have been found, and on the whole they have proved to be closely related to the originally

discovered vector. To what extent does a scheme of differentiation of these specifically transmitted plant viruses on the basis of natural relationships of their vectors also provide a natural classification of the viruses? It has been suggested that viruses represent the ultimate in parasitism (Green, 1935). In any event, their evolution as parasites has been proceeding for a very long time. It seems not unreasonable therefore to consider the probability that, during ages of transmission by insects, certain virus species, genera, etc. differentiated as their insect vectors differentiated into species, genera, and so forth.

There may be plant viruses with no animal vectors, and there are many plant viruses belonging in a group where each has many vectors. In these groups the vector relationships would not provide a satisfactory means of virus differentiation. However, it is possible that these two groups of viruses are natural, viruses within each being more closely related to others in the group than to viruses outside the group. Here, criteria other than vector relationships are necessary for the purpose of differentiation and classification within the groups.

In arriving at an estimate of the value of the vector relationships in differentiating and classifying plant viruses, the use of all other criteria of virus classification, namely, symptoms, protection tests, serology, physical and chemical properties, plant hosts, etc., will be necessary. However, it would seem unwise, on the basis of present knowledge, to hold that relationships indicated by symptoms alone were more real or natural than those indicated by vector relationships alone. Plant pathogens other than viruses cannot be classified into a natural system on the basis of the phylogeny of their vectors but neither can they be so classified on the basis of the symptomatic classification of the diseases they cause. Virus strains may cause a variety of symptoms in one host, and the same strain may cause different symptoms in different hosts. These facts suggest the danger in emphasizing too strongly the use of symptoms as a means of obtaining a natural classification of plant viruses.

SUMMARY

Two hitherto unknown viruses were discovered, the symptoms they produce on crimson clover are described, their vector relationships are determined, and their classification is suggested.

A new genus, *Aureogenus*, is proposed for the potato yellow-dwarf viruses and for the two new viruses. It is proposed that the New York potato yellow-dwarf virus be named *Aureogenus vastans* var. *vulgare* n. comb., the New Jersey potato yellow-dwarf virus *A. vastans* var. *agalliae* n. comb., one of the new viruses *A. magnivena* n. sp., and the other new virus *A. clavifolium* n. sp.

The virus *Aureogenus vastans* var. *vulgare* was transmitted by *Aceratagallia sanguinolenta* (Provancher), *Aceratagallia lyrata* (Baker), *Aceratagallia obscura* Oman, and *Aceratagallia curvata* Oman. It was not transmitted by *Agallia constricta* Van Duzee. It was transmitted by only 2 of 465 *Agallia quadripunctata* (Provancher), and possibly by 1 of 220 *Agalliopsis novella* (Say).

The virus *Aureogenus vastans* var. *agalliae* was transmitted by *Agallia constricta* and *Agallia quadripunctata*, but not by *Aceratagallia sanguinolenta*. It is possible that 1 of 241 *Agalliopsis novella* transmitted this virus.

Aureogenus magnivena was transmitted by *Agallia constricta*, *Agallia quadripunctata*, and *Agalliopsis novella*, not by *Aceratagallia sanguinolenta*.

Aureogenus clavifolium was transmitted by *Agalliopsis novella*, not by *Aceratagallia sanguinolenta* nor by *Agallia constricta* nor by *Agallia quadripunctata*.

There is evidence that the cases of rare transmission are real and are not due to contamination.

Protection tests are described, indicating the relationship of the NY and NJ potato yellow-dwarf viruses. Evidence on the relationships of *Aureogenus magnivena* and *Aureogenus clavifolium* is presented.

LITERATURE CITED

Barrus, M. F., and C. C. Chupp. 1922. Yellow dwarf of potatoes. *Phytopathology* **12**: 123–132.

Black, L. M. 1937. A study of potato yellow dwarf in New York. *New York (Cornell) Agric. Exper. Sta. Mem.* 209: 1–23.

—— 1940. Strains of potato yellow-dwarf virus. *Am. Jour. Bot.* **27**: 386–392.

—— 1941*a*. Further evidence for multiplication of the aster-yellows virus in the aster leafhopper. *Phytopathology* **31**: 120–135.

—— 1941*b*. Specific transmission of varieties of potato yellow-dwarf virus by related insects. *Am. Potato Jour.* **18**: 231–233.

—— and W. C. Price. 1940. The relationship between viruses of potato calico and alfalfa mosaic. *Phytopathology* **30**: 444–447.

Elze, D. L. 1931. The relation between insect and virus as shown in potato leaf roll, and a classification of

viroses based on this relation. *Phytopathology* **21**: 675–686.

FAWCETT, G. L. 1927. The curly top of sugar beet in the Argentine. *Phytopathology* **17**: 407–408.

FUKUSHI, T. 1935. Multiplication of virus in its insect vector. *Proc. Imp. Acad. Japan* **11**: 301–303.

—— 1939. Retention of virus by its insect vectors through several generations. *Proc. Imp. Acad. Japan* **15**: 142–145.

—— 1940. Further studies on the dwarf disease of rice plant. *Jour. Facul. Agric.*, Hokkaido Imp. Univ. **45** (3): 83–154.

GREEN, R. G. 1935. On the nature of filterable viruses. *Science* **82**: 443–445.

HOLMES, F. O. 1939. Handbook of phytopathogenic viruses. Minneapolis.

—— 1941. A distinctive strain of tobacco-mosaic virus from *Plantago*. *Phytopathology* **31**: 1089–1098.

JOHNSON, F. 1942. The complex nature of white-clover mosaic. *Phytopathology* **32**: 103–116.

JOHNSON, J., AND ISMÉ A. HOGGAN. 1935. A descriptive key for plant viruses. *Phytopathology* **25**: 328–343.

KUNKEL, L. O. 1932. Celery yellows of California not identical with the aster yellows of New York. *Contrib. Boyce Thompson Inst.* **4**: 405–414.

—— 1935. Possibilities in plant virus classification. *Bot. Rev.* **1**: 1–17.

—— 1937. Effect of heat on ability of *Cicadula sexnotata* (Fall.) to transmit aster yellows. *Am. Jour. Bot.* **24**: 316–327.

—— 1941. Heat cure of aster yellows in periwinkles. *Am. Jour. Bot.* **28**: 761–769.

OMAN, P. W. 1933. A classification of North American agallian leafhoppers. *U. S. Dept. Agric. Tech. Bull.* 372: 1–94.

—— 1934. South American leafhoppers of the genus *Agalliana*. *Rev. Entomol.* [Rio de Janeiro, Brazil]: **4**(3): 333–340.

SEVERIN, H. H. P. 1934. Experiments with the aster-yellows virus from several states. *Hilgardia* **8**: 305–325.

STOREY, H. H. 1931. The bearing of insect-vectors on the differentiation and classification of plant viruses. *Deux. Cong. Internat. Pathol. Comp.*, pp. 471–479.

—— AND A. P. D. MCCLEAN. 1930. The transmission of streak disease between maize, sugar cane and wild grasses. *Ann. Appl. Biol.* **17**: 691–719.

APPENDIX

Once during the course of the experiments dealt with in tables 1 and 2, the stock non-viruliferous culture of *Agallia constricta* became contaminated with the NJ virus. This contamination may have taken place by means of viruliferous *A. constricta* entering the greenhouse during the summer and infecting some of the clover plants used for rearing the insects. Each year a few adults of *A. constricta* have been found free in the greenhouse, but only during the summer months, a fact that indicates they have entered the greenhouse from the field. Before the contamination of the stock culture was detected, the insects had been employed in experiment 5. Of the 50 individuals of *A. constricta* in experiment 5 that had fed on crimson clover containing NY virus, one infected one test plant with the NJ virus, and of the 47 that had fed on crimson clover containing the BV virus, one insect infected two plants with the NJ virus. That the contamination came from the stock culture of insects was clearly indicated by the failure of any of the test plants fed on by *A. novella* to develop New Jersey yellow-dwarf and by the appearance of the disease among the plants used to maintain the stock culture of insects. The data given for transmission of NJ virus by *A. constricta* in experiment 5, table 1, therefore were obtained by subtracting 1 from both the numerator and denominator of the actual fraction (18/48) obtained, 1 representing the calculated amount of contamination. In the same way the data given for the transmission of NJ virus by *A. constricta* in experiment 5, table 2, have been corrected for the small amount of contamination indicated above.

In the first experiment involving the CL virus, experiment 7, the *Agalliopsis novella* insects that had fed on crimson clover containing the CL virus were the only ones to transmit it, but this was not the case in any of the subsequent three experiments. In experiments 8, 9, and 10, 29 of the 368 *A. novella* fed on crimson clover containing NY, NJ, or BV virus infected 34 out of 927 test plants with CL virus. This, together with the fact that no contaminations with the CL virus occurred among the test plants fed on by *Aceratagallia sanguinolenta* or *Agallia constricta*, indicated that the stock culture of *A. novella* was contaminated with CL virus. The data in tables 1 and 2, experiments 8, 9, and 10, dealing with CL virus transmission by *A. novella*, have been corrected for the contamination in the manner described above. Incidentally, this contamination revealed that a single *A. novella* could infect one test plant with CL virus and the next with BV virus.

SOME TABLES TO FACILITATE COMPUTATION OF ELEMENTS OF ECLIPSING BINARIES

ZDENĚK KOPAL

Harvard College Observatory

ABSTRACT

Tables are provided for a class of functions $I^m_{\beta,\gamma}$ defined by equation (1) below. These functions have recently been found essential to the process of improving a preliminary set of elements of an eclipsing system by least-squares corrections, and to the evaluation of photometric effects of rotational and tidal distortion of the eclipsing component. In accordance with the present-day policies of table-construction and the requirements of war-time economy, compressed four-figure tables of these functions are given for ten values of the independent variable, along with the corresponding modified second tabular differences which should facilitate interpolation. Should it ever become desirable to have the I-functions subtabulated to smaller intervals, intermediate entries can be readily extracted from data given presently without any loss of accuracy.

THE aim of the present paper is to provide suitable tables of a family of functions $I^m_{\beta,\gamma}$, defined by

$$\pi r_2{}^q I^m_{\beta,\gamma} = \int_{\delta - r_2}^{c} [r_2{}^2 - (\delta - x)^2]^{\beta/2} \times [2\delta(s - x)]^{\gamma/2}[\delta - x]^m dx, \quad (1)$$

where r_1 = fractional radius of the eclipsed component,

r_2 = fractional radius of the eclipsing component,

δ = apparent separation of centers of the two stars (the radius of their relative orbit being taken as the unity of length),

$$s = \frac{r_1{}^2 - r_2{}^2 + \delta^2}{2\delta},$$

and $c = s$ or $\delta + r_2$ as to whether the eclipse is partial or annular. The superscript and subscripts of the $I^m_{\beta,\gamma}$'s are integers and $q = 1 + \beta + \gamma + m$.

The functions $I^m_{\beta,\gamma}$ have recently been found to be of twofold application to the process of determining the photometric elements of eclipsing variables from an analysis of their light curves. First, if an adjustment of elements of a system consisting of spherical stars is being made by the method of least squares by means of equations of condition of the form

$$(O - C) = -f\Delta L_1 - L_1\Delta f,$$

$$\Delta f = \frac{\partial f}{\partial r_1}\Delta r_1 + \frac{\partial f}{\partial r_2}\Delta r_2 + \frac{\partial f}{\partial \delta}\Delta\delta + \cdots \quad (2)$$

where f denotes the fractional loss of light within minimum; L_1, the fractional luminosity of the eclipsed star; and the Δ's stand for the respective differential corrections, then [1]

$$\frac{\partial f^U}{\partial r_2} = 2\frac{r_2}{r_1{}^2} I^0_{-1,0}, \qquad \frac{\partial f^U}{\partial \delta} = -2\frac{r_2}{r_1{}^2} I^1_{-1,0}, \quad (3.0)$$

if the eclipsed star is uniformly bright, and

$$\frac{\partial f^D}{\partial r_2} = 3\frac{r_2{}^2}{r_1{}^3} I^0_{-1,1}, \qquad \frac{\partial f^D}{\partial \delta} = -3\frac{r_2{}^2}{r_1{}^3} I^1_{-1,1}, \quad (3.1)$$

if it is completely darkened at limb. Values of the remaining differential quotients follow, furthermore, simply from those already obtained.[2]

Secondly, if the eclipsing component is no longer spherical but gets flattened at poles because of its own rotation and distorted, in addition, by the tidal pull of its companion, the photometric effects of the resulting deformation of its shadow-cylinder upon fractional loss of light within

---

[1] *Cf.* Kopal, *Proc. Am. Philos. Soc.* **86** (3): 342, 1943.

[2] Some tables to facilitate computation of coefficients of equation (2) for limb-darkened stars were previously constructed by Wyse and Kron (*Lick Observ. Bull.* 496, 1939). The reader may indeed verify that the functions A and B studied by Wyse and Kron are particular cases of the $I^m_{\beta,\gamma}$'s to be discussed presently (eq. 6), namely

$$A = \frac{\pi}{\sqrt{2}} I^0_{-1,1} \quad \text{and} \quad B = \frac{\pi}{\sqrt{2}} I^1_{-1,1}$$

irrespective of whether the eclipse is partial or annular, or whether it is an occultation or a transit. Furthermore, Wyse and Kron's parameter $c = -\mu$ of our equation (5). Wyse and Kron evaluated the integrals A and B by approximate methods to three decimal figures and tabulated in terms of c. Such tables are, however, ill suited for interpolation and should be superseded by more accurate tables given presently.

minima turn out to be expressible as [3]

$$\begin{aligned}&-\frac{1}{3}v\{3n_1^2I^0_{1,0}+3n_2^2I^2_{-1,0}-I^0_{-1,0}\}\\&+w_2\{3\delta^2I^2_{-1,0}-I^0_{-1,0}\}\\&+w_3\{5\delta^3I^3_{-1,0}-3\delta I^1_{-1,0}\}\\&+\frac{1}{4}w_4\{35\delta^4I^4_{-1,0}-30\delta^2I^2_{-1,0}+3I^0_{-1,0}\}+\cdots\end{aligned}\tag{4.0}$$

if the star undergoing eclipse appears uniformly bright, and

$$\begin{aligned}&-\frac{1}{2}v\{3n_1^2I^0_{1,1}+3n_2^2I^2_{-1,1}-I^0_{-1,1}\}\\&+\frac{3}{2}w_2\{3\delta^2I^2_{-1,1}-I^0_{-1,1}\}\\&+\frac{3}{2}w_3\{5\delta^3I^3_{-1,1}-3\delta I^1_{-1,1}\}\\&+\frac{3}{8}w_4\{35\delta^4I^4_{-1,1}-30\delta^2I^2_{-1,1}+3I^0_{-1,1}\}+\cdots\end{aligned}\tag{4.1}$$

in case of complete darkening. In these expressions v stands for the oblateness of the eclipsing star; the w_j's are contributions to its equatorial ellipticity owing to the jth tidal harmonic distortion; and

$$\delta^2 = 1 - \cos^2\psi \sin^2 i,$$

while

$$\delta n_1 = \sin\psi \sin i$$

and

$$\delta n_2 = -\cos\psi \sin i \cos i,$$

where, as usual, ψ denotes the phase-angle and i, the inclination of orbital plane to the celestial sphere. Equations (4.0) and (4.1) are complete as far as the disturbing action of the eclipsed star can be regarded as being due to a mass-point.

All functions of the form $I^m_{\beta,\gamma}$ which occur in equations (3) and (4) have recently been evaluated by the writer in a finite number of terms.[4] Those with $\gamma = 0$ or an even integer are found expressible in terms of circular and algebraic functions, while if γ is odd, the $I^m_{\beta,\gamma}$'s involve complete elliptic integrals of the first and second kind. Their complexity increases rapidly with increasing superscript, and their explicit forms are in general too long to be quoted here. They all are by definition positive dimensionless quantities. Those of the form $I^m_{\beta,0}$, valid for uniformly bright stars, can be made to depend on a single variable

$$\mu = \frac{\delta - s}{r_2} = \frac{r_2^2 - r_1^2 + \delta^2}{\delta r_2},\tag{5}$$

which immediately suggests itself as an argument of tabulation for any numerical work. Functions of the form $I^m_{\beta,1}$, relevant to completely darkened stars, depend in principle on two independent variables, but in such a way that

$$I^m_{\beta,1} = \sqrt{\frac{\delta}{r_2}}\, J^m_{\beta,1},\tag{6}$$

where $J^m_{\beta,\gamma}$ is again a function of μ only. Since the whole complexity of the original function is evidently stored in $J^m_{\beta,\gamma}$, it appears expedient in the numerical work to limit ourselves to the construction of uni-variate tables of J as a function of μ rather than to resort to extensive and laborious bi-variate tables.

A preliminary survey of the ground reveals μ itself to be ill suited to serve directly as the argument of tabulation. We notice, for example, that

$$\lim_{\mu\to 1}\left\{\frac{\partial}{\partial\mu} I^m_{-1,0}\right\} = \infty,$$

which makes it obvious that in any table with μ as an argument, an interpolation for $I^m_{-1,0}$ in the neighborhood of $\mu = 1$ should be difficult if not impossible. In order to avoid this singularity, we find it convenient to adopt another independent variable α defined by

$$\alpha = 2\sin^{-1}\kappa,\tag{7}$$

where κ is the modulus of the complete elliptic integrals in which (1) can be resolved if γ is odd. This modulus is known to be

$$\kappa^2 = \frac{1}{2}(1-\mu) \quad \text{or} \quad \frac{2}{1-\mu},$$

as to whether the eclipse is partial or annular. Hence

$$\mu = 1 - 2\sin^2\frac{\alpha}{2}, \quad \text{(partial eclipse)}$$

$$\mu = 1 - 2\csc^2\frac{\alpha}{2}. \quad \text{(annular eclipse)}$$

The new angle α is, within the domain considered, a single-valued function of μ for either type of eclipse; and if latter is partial, their relationship reduces to $\alpha = \cos^{-1}\mu$. Irrespective of the type of eclipse, at the moment of first contact $\alpha = 0^0$. If the eclipsing component

[3] Kopal, *Astrophys. Jour.* **94**: 159, 1941, or (more completely) *Proc. Am. Philos. Soc.* **85**: 399, 1942.

[4] Kopal, *Proc. Am. Philos. Soc.* **85**: 399, 1942 (section 21).

is the larger of the two (occultation), α increases with diminishing light up to a maximum whose location depends on the ratio of radii of the components, and decreases hereafter until it returns to zero at the moment of inner contact (beginning of totality). If, on the other hand, the smaller star is in front (transit), α increases from 0° to 180° at the moment of internal tangency, and decreases during annular phase until it eventually vanishes at the moment when both disks happen to appear concentric ($\delta = 0$).

Before proceeding to give the actual tables, a few considerations concerning the general policy of table-construction may be in order. The determination of elements of eclipsing binaries has more than any other procedure in astronomical use to rely on extensive use of tabulated functions. A precedent has been set in the pioneer investigations by Russell and Shapley [5] to tabulate the required functions at intervals so chosen that an interpolation for intermediate entries could be performed mentally, using first differences only. This was certainly convenient from the computer's point of view and quite feasible as long as three-figure accuracy was considered satisfactory. It is, however, questionable nowadays as to how long this policy could be maintained in view of increasing demands for accuracy on one side, and of increasing costs of computing and publication on the other. A general trend in recent decades in all fields calling for extensive use of tables has been to publish tables in compressed form—that is, to give fewer entries, computed generally to a large number of decimals, along with certain characteristics which should facilitate interpolation for intermediate values—and it appears advisable to extend a similar policy also to the eclipsing orbit work. Compressed tables may be found perhaps somewhat less convenient to their users; but the latter have to be content if the interpolation for a particular entry required will take them as many minutes as it might take hours to compute it *ab initio*.

As an example of compressed tables, values of $I^m_{\beta,\gamma}$ (or $J^m_{\beta,\gamma}$) will be given in the present paper for ten values of α, spaced at intervals of 20°, to four decimal figures. Three decimals would no longer be adequate, while more digits than four might be superfluous for a long time to come. All entries given in the tables were originally computed to five decimals and subsequently rounded off; the errors of tabulated values should therefore not exceed half a unit of the last place.[6] The original five-figure values were differenced mechanically on the National Accounting Machine in a technique described by Comrie [7] as a safeguard against accidental errors as well as the first step toward setting up the interpolation coefficients.

In order to facilitate interpolation for intermediate entries, the present tables contain,[8] in addition to values of the original functions, also the modified second tabular differences M'' defined by

$$M'' = \Delta'' - 0.184\Delta^{IV} + \cdots$$

By means of this device, known as "throw-back," we can namely account, very approximately, for effects of the fourth or higher (even) differences without introducing the corresponding interpolation coefficients. For the actual interpolation we adopt the Everett formula of the form

$$I_n = (1 - n)I_0 + nI_1 + M_0''E_0'' + M_1''E_1'' + \cdots \quad (0 \leqslant n \leqslant 1) \quad (8)$$

as the most convenient one. As is well known, the first four terms of an ordinary Everett formula take account of effects of the third differences as well. Introducing modified second differences, we can, furthermore, account for effects of the fourth differences in the present case without committing any appreciable error. The use of the Everett formula in the above form is therefore sensibly equivalent to that of any other central-difference formula including differences up to the fourth, and its advantage is that only one coefficient (M'') needs to be tabulated for every value of the argument.[9] The second-

[5] Russell and Shapley, *Astrophys. Jour.* **35**: 315, 1912, and **36**: 54, 239, 385, 1913.

[6] A determination of all values of the present three tables did not require the same amount of labor. While for most entries in tables 1 and 2 six-figure values of circular or elliptic functions were as a rule sufficient, the computation of entries in the lower right corner of table 3, resulting as a difference of two large numbers, taxed to capacity the keyboard of a ten-place computing machine. The ten-figure values of the elliptic integrals and of trigonometric functions were taken from vol. 2 of Legendre's *Traité des fonctions elliptiques* and from Andoyer's *Nouvelles tables trigonométriques*, respectively.

[7] Comrie. Scientific application of the National Accounting Machine, etc. *Suppl.* to *Jour. Royal Statistical Soc.* **3**: 87, 1936.

[8] Except for $I^0_{-1,0}$ where the interpolation is strictly linear.

[9] Since, furthermore, both E_0'' and E_1'' are always numerically smaller than one-tenth, it is legitimate presently to retain only three decimals in M''.

TABLE 1

UNIFORM STARS, PARTIAL ECLIPSE

α	$I^0_{-1,0}$	$I^1_{-1,0}$	M''	$I^2_{-1,0}$	M''	$I^3_{-1,0}$	M''	$I^4_{-1,0}$	M''	$I^0_{-1,0}$	M''
0°	0.0000	0.0000	−0.000	0.0000	0.000	0.0000	0.000	0.0000	0.000	0.0000	0.000
20	0.1111	0.1089	−0.013	0.1067	−0.026	0.1046	−0.038	0.1026	−0.050	0.0044	+0.026
40	0.2222	0.2046	−0.025	0.1895	−0.040	0.1764	−0.048	0.1651	−0.048	0.0327	+0.040
60	0.3333	0.2757	−0.034	0.2356	−0.035	0.2068	−0.019	0.1853	−0.014	0.0978	+0.035
80	0.4444	0.3135	−0.039	0.2494	−0.014	0.2153	−0.006	0.1875	+0.002	0.1950	+0.014
100	0.5556	0.3135	−0.039	0.2506	+0.014	0.2153	−0.006	0.1875	−0.002	0.3050	−0.014
120	0.6667	0.2757	−0.034	0.2644	+0.035	0.2068	−0.019	0.1897	+0.014	0.4023	−0.035
140	0.7778	0.2046	−0.025	0.3105	+0.040	0.1764	−0.048	0.2099	+0.048	0.4673	−0.040
160	0.8889	0.1089	−0.013	0.3933	+0.026	0.1046	−0.038	0.2724	+0.050	0.4956	−0.026
180	1.0000	0.0000	0.000	0.5000	0.000	0.0000	0.000	0.3750	0.000	0.5000	0.000

TABLE 2

DARKENED STARS, PARTIAL ECLIPSE

α	$J^0_{-1,1}$	M''	$J^1_{-1,1}$	M''	$J^2_{-1,1}$	M''	$J^3_{-1,1}$	M''	$J^4_{-1,1}$	M''	$J^0_{1,1}$	M''
0°	0.0000	+0.061	0.0000	+0.063	0.0000	+0.064	0.0000	+0.065	0.0000	+0.066	0.0000	−0.002
20	0.0303	+0.059	0.0298	+0.055	0.0294	+0.050	0.0289	+0.046	0.0285	+0.041	0.0009	+0.009
40	0.1188	+0.052	0.1117	+0.032	0.1055	+0.018	0.1000	+0.007	0.0951	0.000	0.0133	+0.034
60	0.2587	+0.041	0.2252	+0.004	0.2002	−0.012	0.1811	−0.015	0.1661	−0.015	0.0584	+0.052
80	0.4388	+0.026	0.3423	−0.025	0.2871	−0.022	0.2512	−0.014	0.2259	−0.010	0.1517	+0.048
100	0.6440	+0.007	0.4356	−0.046	0.3562	−0.008	0.3078	−0.011	0.2750	−0.010	0.2879	+0.017
120	0.8555	−0.016	0.4858	−0.051	0.4162	+0.005	0.3503	−0.025	0.3149	−0.004	0.4394	−0.023
140	1.0509	−0.041	0.4880	−0.038	0.4810	+0.011	0.3680	−0.040	0.3523	+0.013	0.5699	−0.052
160	1.2033	−0.074	0.4552	−0.014	0.5509	−0.021	0.3524	−0.022	0.3978	+0.003	0.6524	−0.058
180	1.2732	−0.171	0.4244	+0.097	0.5942	−0.120	0.3274	+0.084	0.4325	−0.100	0.6791	−0.054

difference coefficients E_0'' and E_1'' of the Everett formula have been extensively tabulated in terms of n by several writers.[10] Tables sufficient for our purposes were published in the *Nautical Almanac* for 1937 and are therefore accessible to every astronomer.[11]

With the aid of data given in the preceding tables 1 and 2 and equation (8) we can interpolate for any intermediate entry with an error not in excess of one unit at the last place. The following example should illustrate the procedure. Let us consider eclipse of the smaller star by the larger (occultation) in a system characterized by the ratio of radii k, say, 0.784, and let it be required to find the value of $I^1_{-1,1}$ at the moment when the geometrical depth of the eclipse $p = -0.531$. From table 2 we get

$$J^1_{-1,1} = 0.1117(1 - n) + 0.2252n + 0.032E_0'' + 0.004E_1''$$

while the argument n can be found as follows. Since the eclipse considered is an occultation,

$$\mu = 1 - \frac{k^2(1 - p^2)}{2(1 + kp)} = 0.6203,$$

and $\alpha = \cos^{-1}\mu = 51°.66$. Hence $n = 0.5830$, and table 21 (p. 786) of the *Nautical Almanac* for 1937 yields $E_0'' = -0.057$ and $E_1'' = -0.064$. Inserting these figures in the above formula, we obtain $J^1_{-1,1} = 0.1758$, while a direct evaluation of the respective integral yields 0.17581 correctly to five decimals. In order to obtain $I^1_{-1,1}$ we still have to multiply

[10] Coefficients of the Everett formula have been tabulated at intervals of 0.001 in n by E. Chappell (*A Table to Facilitate Interpolation by the Formulae of Gauss, Bessell, and Everett*, printed and published by the author, 41 Westcombe Park Road, London S.E. 3), and by A. J. Thompson (*Table of the Coefficients of Everett's Central-Difference Interpolation Formula*, Cambridge Univ. Press, 1921). Remembering that the second-difference Everett coefficients are identical with outer coefficients of the Lagrangian four-point formula, the reader will find, for instance, another table of E_0'' and E_1'' computed at intervals of 0.001 in *The Kelley Statistical Tables* (Macmillan, 1938), under the heading of cubic interpolation coefficients.

[11] Tables XX (p. 785), XXI (p. 786), and XXVI (p. 800–801). Reprinted by L. J. Comrie under the title *Interpolation and Allied Tables* (H. M. Stationery Office, London, 1936).

TABLE 3

DARKENED STARS, ANNULAR ECLIPSE

α	$J^0_{-1,1}$	M''	$J^1_{-1,1}$	M''	$J^2_{-1,1}$	M''	$J^3_{-1,1}$	M''	$J^4_{-1,1}$	M''	$J^0_{1,1}$	M''
180	1.2732	+0.168	0.4244	−0.108	0.5942	+0.116	0.3274	−0.081	0.4325	+0.100	0.6791	+0.051
160	1.3448	+0.089	0.3916	−0.007	0.6405	+0.041	0.2991	−0.005	0.4712	+0.028	0.7043	+0.049
140	1.5168	+0.083	0.3396	−0.002	0.7382	+0.038	0.2571	+0.001	0.5483	+0.027	0.7784	+0.046
120	1.7805	+0.120	0.2846	+0.003	0.8786	+0.057	0.2144	+0.003	0.6559	+0.042	0.9019	+0.063
100	2.1699	+0.187	0.2318	+0.003	1.0787	+0.092	0.1741	+0.003	0.8075	+0.070	1.0912	+0.096
80	2.7596	+0.379	0.1816	+0.003	1.3768	+0.188	0.1364	+0.002	1.0327	+0.147	1.3828	+0.153
60	3.7369	+0.541	0.1340	+0.002	1.8673	+0.267	0.1004	+0.002	1.4003	+0.209	1.8697	+0.453
40	5.6726	—	0.0881	+0.002	2.8360	—	0.0661	+0.001	2.1268	—	2.8367	—
20	11.4302	—	0.0439	—	5.7152	—	0.0329	—	4.5081	—	5.7150	—

by $\sqrt{\delta/r_2}$. Since for an occultation $\delta/r_2 = 1 + kp$, $\sqrt{\delta/r_2} = 0.7640$ and thus, finally $I^1_{-1,1} = 0.1343$.

The use of table 3, appropriate for annular eclipses of darkened stars,[12] requires little more caution. If, namely, $\alpha \to 0°$, the integrals $J^m_{\beta,\gamma}$ cease to be regular and tend to $\pm\infty$ as to whether their superscript is even or odd. This need not disturb the reader unduly; for during annular phase α can vanish only if $\delta = 0$; and it can be proved that, for $m > 0$,

$$\lim_{\delta \to 0} \delta^m I^m_{\beta,\gamma} = 0,$$

while if $m = 0$,

$$\pi I^0_{\beta,\gamma} = B\left(\frac{1}{2}, \frac{1}{2}\beta + 1\right) k^{-\gamma}(1 - k^2)^{\gamma/2},$$

where B stands for complete beta-function. Values of $J^m_{\beta,\gamma}$ within, say, the interval $0° < \alpha < 60°$ will be but seldom called for in practice. Should they be required, the computer may try many things (for instance, to interpolate reciprocals); but he may in the long run find it expedient to resort to the original formulae and evaluate the respective entries individually.

[12] During annular eclipse of a uniformly bright star

$$I^{2n}_{-1,0} = \binom{n - \frac{1}{2}}{n} \quad \text{or} \quad I^{2n+1}_{-1,0} = 0,$$

according to whether the superscript is even or odd.

SYMPOSIUM

on

WARTIME ADVANCES IN MEDICINE

and the

R. A. F. PENROSE, JR., MEMORIAL LECTURE

Papers read before

The American Philosophical Society

Annual General Meeting

April 20–22, 1944

PHILADELPHIA

THE AMERICAN PHILOSOPHICAL SOCIETY

INDEPENDENCE SQUARE

1944

CONTENTS

SYMPOSIUM ON WARTIME ADVANCES IN MEDICINE

The Water Requirements of Castaways 151
JAMES L. GAMBLE, Harvard Medical School.

Blood, Blood Derivatives, and Blood Substitutes 159
EDWIN J. COHN, Harvard Medical School.

The Present Status of Penicillin in the Treatment of Infections 174
CHESTER S. KEEFER, Boston University School of Medicine.

Recent Advances in the Treatment of Wounds 177
ALLEN O. WHIPPLE, Columbia University.

The Role of Immunization in Wartime Advances in Medicine 182
R. E. DYER, U. S. Public Health Service.

Human Problems in Military Aviation 189
DETLEV W. BRONK, University of Pennsylvania.

Environmental Protection 196
GEORGES F. DORIOT, Quartermaster Corps, War Department.

Neuropsychiatry 204
EDWARD A. STRECKER, University of Pennsylvania.

Trends in the Study and Control of Infectious Diseases 208
RENÉ J. DUBOS, Harvard Medical School.

R. A. F. PENROSE, JR., MEMORIAL LECTURE

The International Labor Organization Lessons for Twenty-five Years 214
HAROLD B. BUTLER, Minister from Great Britain to the United States.

THE WATER REQUIREMENTS OF CASTAWAYS[1]

JAMES L. GAMBLE

Professor of Pediatrics, Harvard Medical School

(Read April 20, 1944, in the Symposium on Wartime Advances in Medicine)

IN this country and in England there are many men at work in laboratories in behalf of men adrift. I will attempt only a brief presentation of some of the results of a study of the problem of provisioning castaways carried out at the Massachusetts General Hospital in collaboration with Dr. Allan M. Butler. Our findings will perhaps serve to bring before you some of the features of this field for inquiry which the war has produced. These investigations were undertaken before the merciful invention of methods of obtaining water from sea water. Their premise is, therefore, a restricted availability of water. The necessity for limiting weight in providing water and food was so severe that it produced an initial question: should not all of the weight allowance per man per day be devoted to water, on the ground that survival for fasting is so much longer than for thirsting? So it was evident that we should begin our studies by attempting to define the water requirement during fasting.

THE WATER REQUIREMENT DURING FASTING

The components of the water exchange when there is no food intake are indicated by the diagram (fig. 1) There are two items of expenditure, water which leaves the body as water vapor by way of the lungs and skin, the so-called insensible loss, and water removed by the kidneys. This outgo is in small part covered by water produced by oxidation of body protein, fat, and glycogen. Oxidation of body protein incidentally releases intracellular water to an extent which is roughly proportional to the protein-water content of cell fluid. There is also an accompanying loss of extracellular fluid. Since this release of water results from consumption of body substance, it is not a process of dehydration. This is water which, under the circumstances imposed by fasting, is physiologically expendible. Together with water of oxidation, it covers a large part of total water outgo. It we proceed to measure as closely as we can insensible and urine water, water drunk, and water of oxidation, we obtain by difference a measurement of available preformed water within the body.

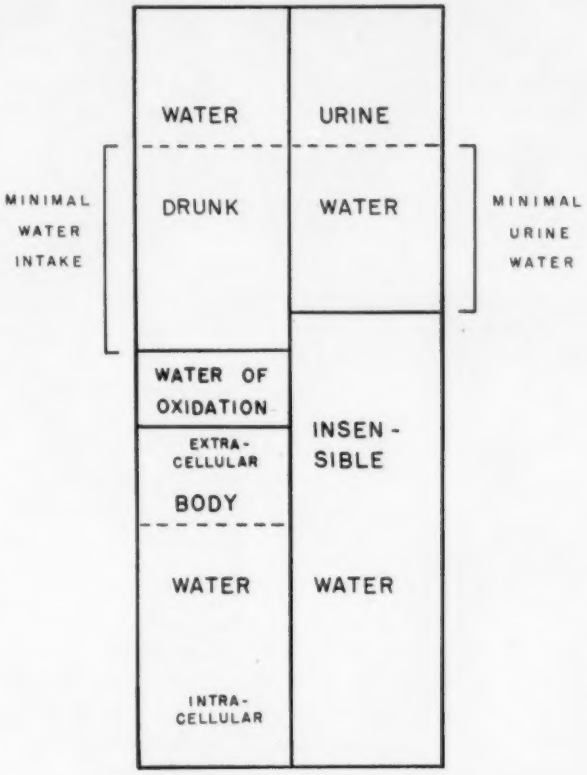


FIG. 1. Water exchange during fasting, with surplus water intake.

Water drunk under guidance of the sensation of thirst usually provides a generous margin over obligatory expenditure, indicated by the broken line, and so leaves a large surplus to be removed

[1] The work described in this paper was done under a contract recommended by the Committee on Medical Research between the Office of Scientific Research and Development and the Massachusetts General Hospital.

Admirable studies of water requirements both at sea and in hot arid regions have been carried out under contracts recommended by C.M.R. between O.S.R.D. and the University of Rochester by a group of investigators working under the direction of Professor Edward F. Adolph.

in the urine. This margin of safety for the water exchange is, of course, in the best physiological tradition, but the castaway cannot afford it. And so we are asked to define obligatory outgo. Since the larger component, the insensible expenditure, has a direct relationship with the metabolism of energy, it should be at its basal value if our subject is at rest. So that what is left for definition is the minimal renal water requirement. If a value can be found for this, the position of the broken line will be established and will produce, on the other side of the diagram, the measurement we are after; the minimal water intake.

The renal water requirement may be approached from a measurement of the freezing point depression of the urine. The few simple steps along the way are shown in table 1. Here

TABLE 1

$\frac{\Delta}{1.86}$ = os M./L.

os M./L. × urine vol. (cc.) = total solutes (milliosmols).

Maximal concentration of urine, 1.4 os M./L.

$\frac{\text{Total solutes}}{1.4}$ = minimal urine volume.

Example:

Urine vol. 1,240 cc. Δ, −1.19°C.

$\frac{1.19}{1.86}$ = 0.64 os M./L.

1240 × 0.64 = 794 milliosmols total solutes.

$\frac{794}{1.4}$ = 566 cc. minimal urine volume.

for example is a twenty-four hour collection of urine of 1,240 cc. We find a freezing point depression of −1.19°C. The physical chemists tell us that if we divide degrees of freezing point depression by 1.86 we obtain osmolar concentration. Multiplying osmolar concentration by cubic centimeters of urine, defines the total output of solutes as milliosmoles. That is, we have an overall measurement of the kidneys' excretory assignment in terms of concentration equivalents. Dividing this output by a value taken as maximal for the ability of the kidney to concentrate solutes in urine, defines the minimal volume of urine required to accomplish removal of this solute load. We have taken 1.4 osmolar as maximal on the basis that it is the average value found for our subjects over periods of thirsting and over other periods of limited water intake.

To illustrate application of this measurement; the left hand diagram in figure 2 is constructed from average daily values found for the components of the water exchange over a six-day fast. The subject received an abundant water intake and so it may be assumed that body water was not withdrawn beyond the physiologically available quantity. It is also assumed that the insensible water loss is not appreciably above its

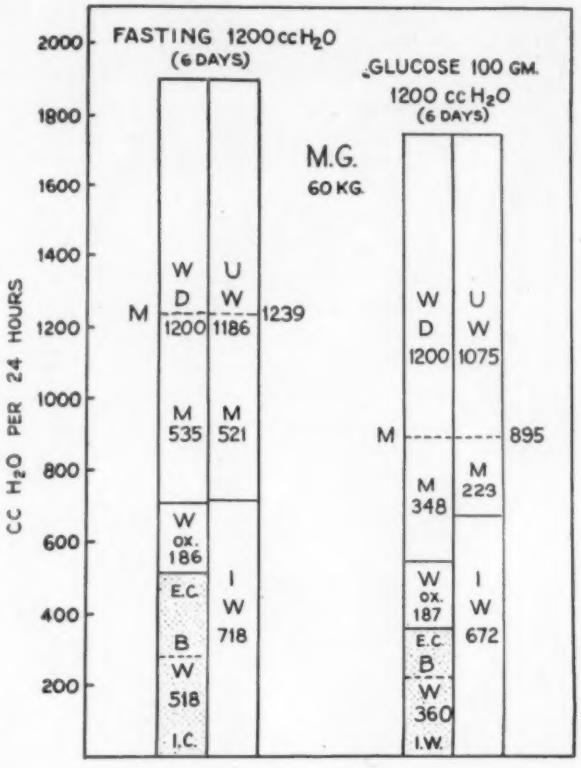


FIG. 2.

basal value. The estimation of minimal urine water, derived from measurement of total solute output, thus completes definition of obligatory expenditures and provides, on the other side of the diagram, a measurement of the water intake requirement which, for this subject, was found to be 535 cc. This is an unusually low value. Measured in this way, the daily water requirement during fasting, when there is little or no physical activity, was found to be, for our other subjects, in the neighborhood of 700 cc.

EFFECTS OF GLUCOSE ON THE WATER EXCHANGE

It was next undertaken to observe effects on the water exchange of providing a small food

intake in the form of glucose. These are shown in the other diagram. Here the subject was given 100 grams of glucose daily along with 1,200 cc. of water. The most conspicuous change is a large reduction of the renal water requirement as compared with the value found during fasting. Since these values for minimal urine water are derived from measurements of total solute output, what glucose does is to reduce greatly the total quantity of substances claiming removal in urine. This would be expected from the protein sparing effect of glucose. The quantity of nitrogenous end products and also of constituents of intracellular fluid presenting for excretion should be reduced to such extent as oxidation of body protein is lowered. On this premise an equivalent reduction of available body water might be expected. Reduction is found but of much less extent than the large lowering of minimal urine water. This circumstance is of great advantage to the castaway. The larger reduction of required water expenditure by the kidney than of available body water produces, as shown in the diagram, a smaller intake requirement than is found for fasting.

The arithmetic of this water gain is shown in table 2. When glucose is given, reduction of the

TABLE 2

EFFECT OF GLUCOSE ON WATER EXCHANGE

Subject M.G.	Minimal Urine Water, cc.	Available Body Water, cc.
Fasting	521	518
100 gm. glucose	223	360
Reduction	298	158

Gain in water exchange 298–158 = 140 cc.

Values average per 24 hrs. for 6-day periods omitting first day.

renal water requirement is larger by 140 cc. than the reduction of available body water. This gain is above the weight of glucose given, which was 100 grams. The significance of this for the castaway is that we are permitted to substitute glucose for a part of the minimal water intake, as defined for fasting, at no cost to the water exchange. So this finding disposes of debate as to whether the castaway's limited intake allowance should be composed entirely of water. According to these measurements, at least 100 cc. of the water requirement for fasting can be replaced by glucose, and all of the physiological benefits of glucose can be gained, without disturbing water balance. Benefit from glucose is very evident. The cheerfulness and the sense of physical effectiveness of our subjects when they were given glucose was in striking contrast with the unhappy lassitude of fasting. Glucose quite clearly contributes to courage which is something of an asset to a castaway.

It turns out that this bit of metabolic legerdemain which permits substitution of a modicum of glucose for water is explained by removal of the ketosis of fasting by glucose. In table 3, ketosis,

TABLE 3

EFFECT OF GLUCOSE ON OUTPUT OF SOLUTES IN URINE

Subject—M.G.	Milliosmols per 24 Hrs.			
	Total Solutes	Organic Acids	NH_4	Org. Ac. $+NH_4$
Fasting	784	133	86	219
100 gm. glucose	362	39	21	60
Reduction	422	94	65	159

159/422 = 0.38

Data from 6-day periods (omitting first day).

in terms of urine solutes, is measured by the increased output of organic acids and of the organic base ammonium, above the values found when glucose is given. The value thus defined for the solutes which derive from the ketosis of fasting was, for this subject, 159 milliosmoles. The absence of these solutes when glucose is given accounts for thirty-eight percent of the total reduction of solute output. In other words, both the antiketogenic and the protein sparing effects of glucose reduce solute output, and thereby the renal water requirement, but only the protein sparing effect relates to the reduction of available body water. This explains the larger reduction of water expenditure than of available water which lowers the intake requirement to an extent which provides just about the right amount of space for the glucose.[2]

SALT WATER EXPERIMENTS

Another item of our inquiry was to examine the suggestion which has so often been put forward that the body may be able to gain, and apply to its expenditures, water from ingested sea water.

[2] It should be noted that the protein sparing and antiketogenic effects of glucose are not unique. They would be gained by an equivalent intake of other food sugars or starches.

This possibility would seem to rest entirely on renal performance. If the kidneys are able to secrete urine at an osmolar concentration above that of sea water, then theoretically water can be gained. Wishing to use scientifically sacred sea water we asked Dr. Alfred Redfield of the Oceanographic Institution at Woods Hole to send us some. Its solute concentration was found to be almost exactly 1.0 osmolar; actually it was 0.97. As we have just seen, the kidney can do better than that. The theoretical gains from a liter of ingested sea water, according to solute concentration in the urine, are shown in table 4.

TABLE 4

Urine Water Requirement for Removal of Solutes of 1.0 Liter of Ingested Sea Water (Solutes in sea water 1.0 os M./L.)

Urine Conc., os M./L.	Water Required, cc.	Water Gained, cc.
1.0	1,000	0
1.1	910	90
1.2	834	166
1.3	770	230
1.4	714	286
1.5	667	334

If urine is secreted at the same concentration as that of sea water there will be no water gain. At a urine concentration of 1.4 osmolar, the solutes from a liter of sea water can be removed in 714 cc. of water and there will be 286 cc. left over, which the body can apply to other expenditures. So at this rate the castaway could cover his daily water requirement by drinking three or four liters of sea water. But this is far beyond gastro-intestinal tolerance for sea water. Ingestion of more than 500 or 600 cc. exerts disturbing effects.

The diagrams in figure 3 present for comparison measurements obtained over a period of thirsting and over another period during which the subject drank 500 cc. of sea water daily. One of our subjects was courageously willing to undergo these quite severe experiments. For the period of thirsting, urine water was 501 cc. An intake of 500 cc. of sea water, if there is no water gain from it, should therefore increase urine water to 1,001 cc. Water expenditure by the kidney is found to be 860 cc. and so a considerable water gain is defined. Under the circumstances of these experiments body water is necessarily withdrawn far above the physiologically available quantity. Available body water for this subject was determined over a period of fasting with an ample water intake and is shown by the broken line across the chart. Water gain should, therefore, be reflected by reduction of body water loss. As may be seen in the diagrams, reduction is found although of less extent than the absolute water gain by the kidneys owing to the circumstance that a part of this gain is dissipated by a small increase of insensible water expenditure over the sea water period.

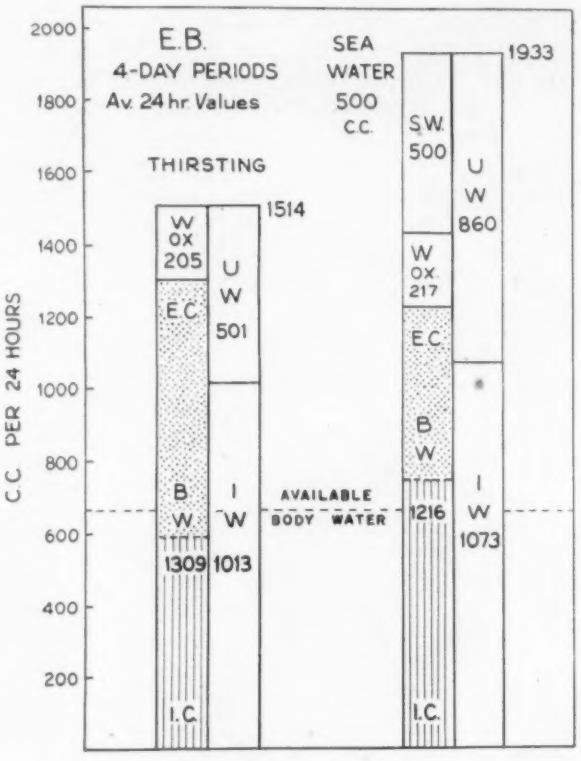


FIG. 3.

Further inspection of the diagrams reveals an additional effect of sea water; a reduction of the loss of extracellular water which is larger than the reduction of the total body water loss, with the result that withdrawal of intracellular water is increased. This partition of body water loss, as regards source, is based on measurements of the accompanying losses of the extracellular ion, sodium, and of the intracellular ion, potassium. It is not surprising that the large intake of sodium which sea water provides should reduce sodium deficit. Extracellular fluid volume within the body is thereby conserved but water for this beyond the direct gain from sea water must be withdrawn from the intracellular compartment and this is shown by an increased removal of potassium in the urine.

Further data for this discussion are presented in table 5. The increase in urine water, 359 cc., with ingestion of 500 cc. of sea water defines a water gain of 141 cc. The theoretical gain, 157 cc., is not quite attained owing to a slightly lower solute concentration over the sea water period. It would of course not be cricket for the kidneys to save water by not removing solutes. The sea

TABLE 5

SEA WATER EXPERIMENT INTAKE 500 CC.

Subj.	Intake	Urine cc.	Solutes		Na Balance, meq.	K Balance, meq.
			osM./L.	Total, m-osM.		
E.B.	Thirsting.....	501	1.46	729	−63	−53
	Sea water 500 cc.	860	1.44	1,235	−49	−73
	Increase.....	359		506		

water being an approximately osmolar solution 500 cc. contains 500 milliosmoles of solutes. As shown by the value found for increase of solute output, 506 milliosmoles, the kidney apparently removed the injested solutes with an excellent accuracy. Turning, however, to the ions which govern water distribution within the body, we find, when sea water is given, a reduction of negagive sodium balance and an increased deficit of potassium. So that, although an appreciable water gain from sea water is clearly demonstrated, there is a surreptitious swapping of electrolytes between sea water and the body fluids of which we are not sure that we approve. Whether or not there is physiological advantage in sustaining extracellular fluid at some cost to intracellular fluid is a troublesome question.

A series of experiments in which sea water supplements were given along with a limited intake of fresh water was carried out. The outcome was essentially the same as for sea water alone.

EXTRA-RENAL WATER EXPENDITURE

The other component of water expenditure will now be briefly considered. Returning to figure 1, the insensible water loss has an independent position in the diagram. It can be dealt with directly. So it will not be necessary to play difficult dominoes to define it. Obviously solicitude over an economical use of water by the kidney will be of little avail to the castaway unless this larger expenditure by way of the lungs and skin is held at somewhere near its basal value. This will require avoidance of physical activity, to such extent as this may be possible, because of the relationship of insensible water outgo to the energy metabolism. Also, in hot weather there is hazard of a large additional vaporization of water in defense of body temperature. So there is need for study of measures which may be expected to prevent or limit this increase of water expenditure. These measures will consist in either direct reduction of environmental warmth or in substituting for the vaporization of body water other processes of heat removal. Various ways of keeping the body cool are simple and evident. They are listed in table 6.

TABLE 6

1. Provision of shade.
   Direct reduction of environmental temperature.
2. Increasing the cooling effect of breeze by removal of clothing.
   Promotion of convection.
3. Periodic immersion in the sea.
   Promotion of conduction.
4. Wetting clothing with sea water.
   Vaporization of sea water in place of body water.

Some of the results of a field study which was undertaken with the purpose of obtaining roughly quantitative information regarding the extent of the water losses and the effectiveness of measures of limiting them under conditions to which castaways would be exposed in hot weather will be briefly presented. These experiments were carried out during the first two weeks of last August in a small cove at the upper end of Cotuit Bay on Cape Cod. Equipment consisted of a rowboat and a small unrigged sailboat with a large and handsome garden umbrella stepped in place of the mast to provide shade. A raft was also used and was provided with a wind screen of stout cloth on all sides which served the purpose of excluding the cooling effect of breeze. The measurements of water loss were obtained over periods of five to eight hours, by the simple procedure of weighing the subjects nude at the beginning and at the end of the period and measuring, in terms of weight, the intake of water and food and the outgo of urine. The subjects drank according to thirst and were given a noontime meal which consisted of fifty grams of a mixture of glucose and fat which resembles butterscotch.

The results of these very simple experiments can easily be seen in the charts (fig. 4). Since sweat was often a large component of the water loss, and since sweat is a sensible loss, it was de-

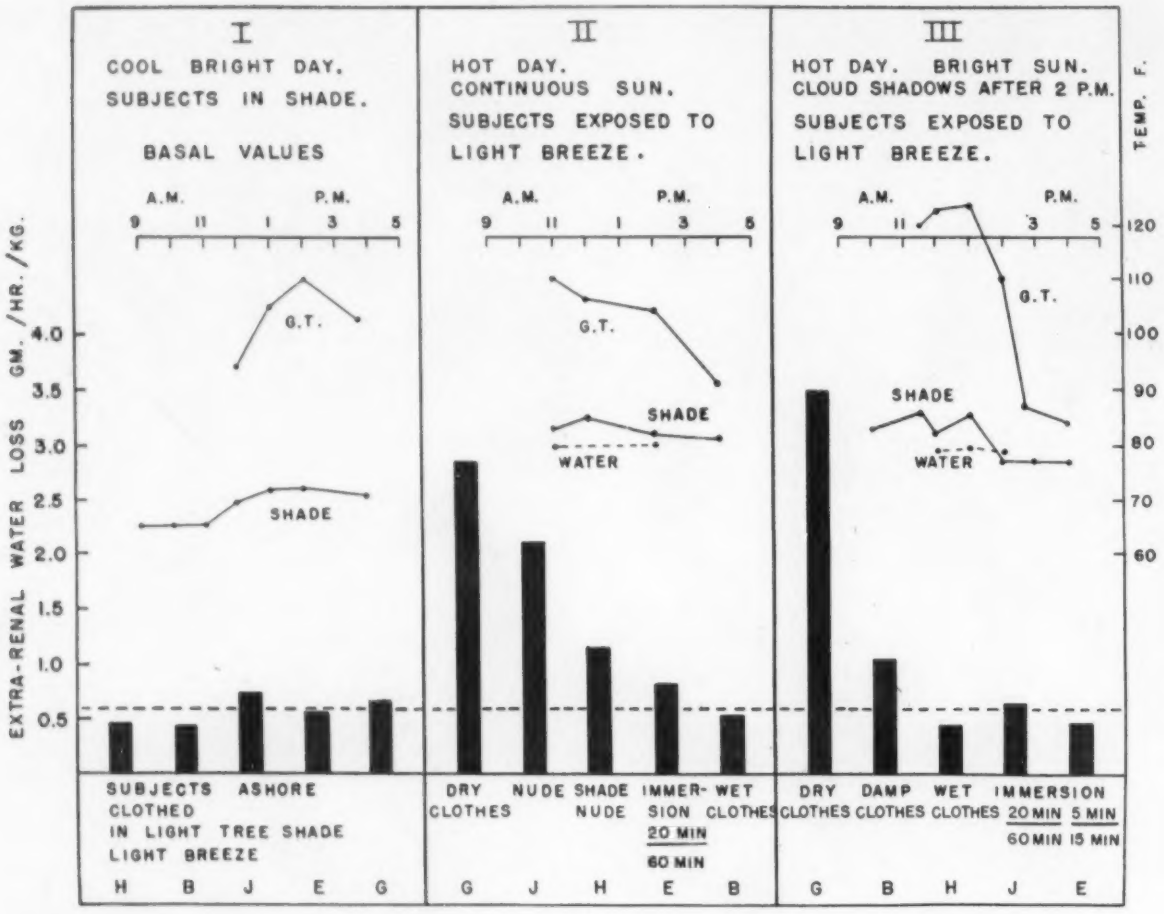


FIG. 4.

cided to call the overall measurement, extra-renal water loss. Expressed as grams per kilogram of body weight per hour, the values found for the individual subjects are shown by the black columns. The environmental circumstances are indicated in the upper part of the charts. Shade temperature was taken under the garden umbrella. The measurements by the so-called Globe Thermometer (G.T.) were taken in the sun and serve to define roughly the increase in environmental warmth above the shade temperature, referable to direct solar radiation.

The measurements in the first section of the chart were obtained over an eight hour period during which the subjects were not on the water. They reclined ashore in the light and fragrant shade of pine trees. They were pleasantly cool and there was no sweating. So these are measurements of the basal rate of insensible water loss. Their average value is 0.6 grams of water per kilogram per hour. This value for irreducible water expenditure is carried across the charts by the broken line.

The data in the next two sections of the chart make clear at a glance both the need for and the effectiveness of measures to limit water expenditure. The losses for subject G, who did nothing to avoid water wastage, were at five to six times the basal rate. Periodic immersion and wet clothing completely prevented loss above the basal rate. The subjects immersed were provided with Kapok life jackets and floated about with a minimum of physical exertion. They were in the water one-third of the time. Immersion to this extent in water at 80° F. produced cooling to the point of slight physical discomfort. They had gooseflesh when they came out but did not shiver. The subjects assigned to wet clothing got into the water with their clothes on and then crawled into the boat and lay there sopping wet. When they became partially dry they did it again. This procedure came to be known as

"dunking," and consistently scored low basals. It was also popular. The subjects were pleasantly and continuously cool. On these days the subjects were in the boats and were exposed to the light breezes which entered the cove. The cooling effect of breeze is, therefore, an undetermined component of these measurements. It is clearly defined by the measurements recorded in the first section of figure 5. Subject H sat in a boat in his bathing trunks and was exposed to breeze. His water expenditure was almost exactly on the basal mark. Subject G, also in his bathing trunks, lay on the raft behind the wind screen and lost water at about five times the basal rate. These two measurements describe very strikingly the beneficence of breeze. The measurements obtained when immersion was placed on a voluntary basis (section IV and V) were interestingly informative. The subjects were instructed to go into the water to such extent as would produce comfortable cooling. On this basis immersion fell far short of defending the basal rate of water loss. Evidently comfort does not define adequate cooling. Subject H

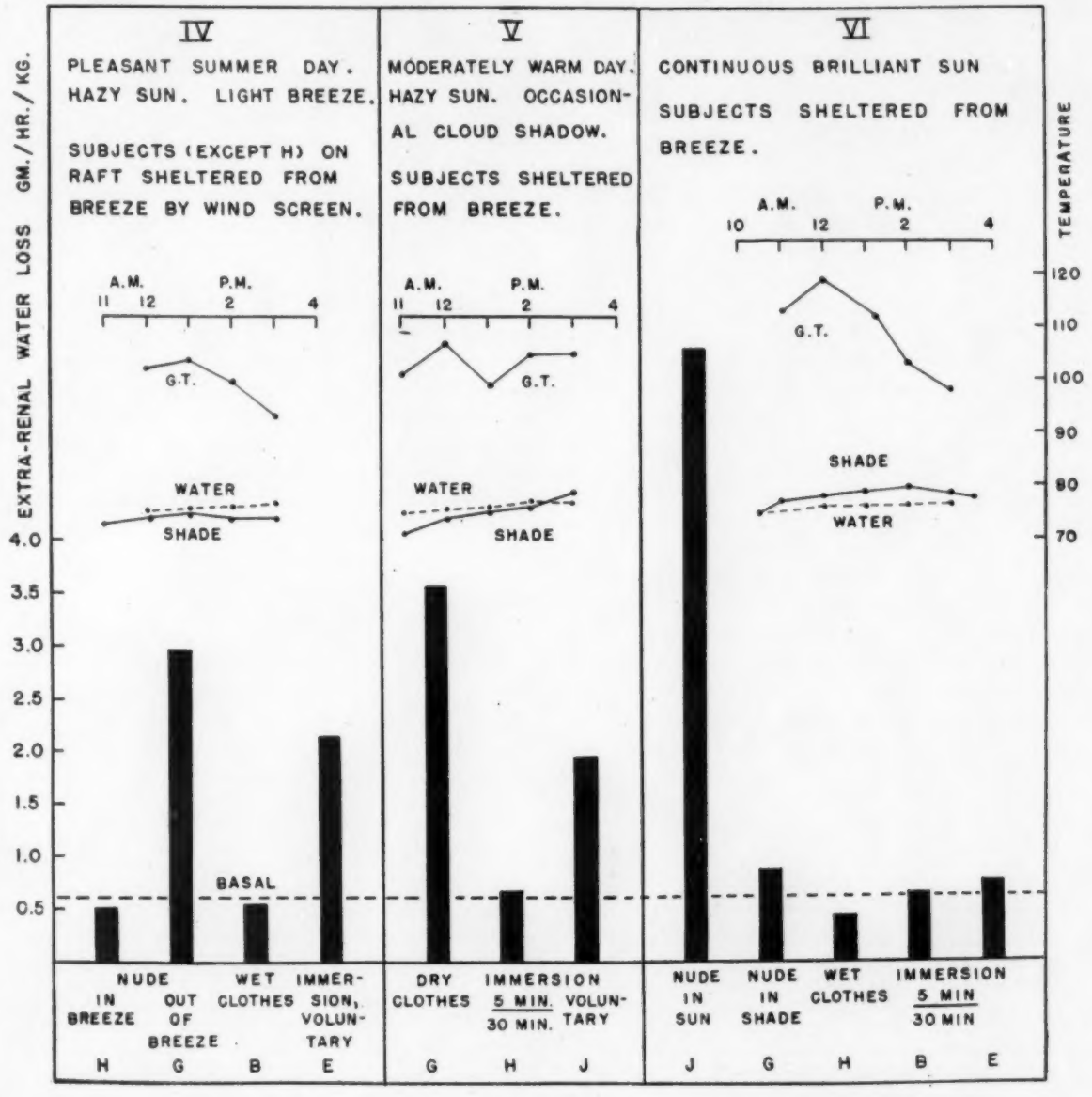


FIG. 5.

who was in the water five minutes of each half hour achieved a basal expenditure. These were overcast days. What unobstructed solar radiation, in the absence of breeze, can do in the way of causing water wastage is shown in the last section of the chart. Subject J, sheltered from breeze and exposed to the sun of a cloudless day of only moderate shade temperature, sweated copiously and lost water at ten times the basal rate. The actual volume of water lost was something more than two liters. He squandered in six hours a quantity of water which should cover his intake requirement over a period of three days. As shown by the measurements from the other subjects on the raft, who were also sheltered from breeze, this dangerously rapid rate of water loss was completely prevented by shade, by periodic immersion, or by wet clothing.

During our outing on Cotuit Bay we were most of all impressed by the protection which breeze provides. The position of the castaway's water balance, if he does nothing about it, will largely depend on the outcome of a contest between the wind and the sun. Fortunately for the castaway, there is nearly always breeze. His greatest hazard is hot windless weather. Then disaster can come very quickly. Provision of shade is clearly commanded and it is very important that equipment for shade should be so designed as not to obstruct breeze. Periodic immersion and wet clothing have a supplementary importance. In shark infested waters one might easily choose wet clothing.

The findings presented in this paper are but fragments of a body of knowledge concerning water requirements under various vicissitudes, which is being gained from the work of many investigators.

BLOOD, BLOOD DERIVATIVES, AND BLOOD SUBSTITUTES*

EDWIN J. COHN

Department of Physical Chemistry, Harvard Medical School

(*Read April 20, 1944, in the Symposium on Wartime Advances in Medicine*)

HUMAN BLOOD

In this war military surgeons have made extensive use of the transfusion of blood or its constituents. No so-called blood substitute that has been considered completely replaces human blood. Whole blood is especially needed when there has been severe blood loss or when major operations are to be performed in which severe blood loss is anticipated. The red cells which are responsible for the respiratory function of the blood, carrying oxygen to the tissues, are also needed to combat the anemia which frequently occurs in convalescence from wounds. The therapeutic value of the red cells is thus related to their physiological functions. The therapeutic value of other constituents of blood may also be related to their physiological functions.

The lability of whole blood made it necessary to prepare for this war by the accumulation of reserves of the more stable components of blood. However, work continues in many laboratories in an attempt to increase the span of life of its labile cellular constituents so as to render them available for military medicine. The Armed Forces have relied thus far on locally available blood donors for whole blood transfusion and on the plasma proteins as blood derivatives which could be prepared in the United States from blood collected by the American Red Cross, processed to a stable state, and accumulated for transport to theatres of operation.

THE PLASMA PROTEINS

Plasma contains components derived from different tissues and serving different functions, many of them unrelated. The specialized molecules which carry out these functions differ in many respects. Thus the immunological functions are, for the most part performed by a small proportion of the proteins known as the gamma globulins. Coagulation of the blood is effected through the transformation of another globulin, prothrombin, into thrombin and its interaction with the most asymmetric protein of the blood stream, fibrinogen, which forms the matrix of the clot. Other globulins are hormones, enzymes, isohemagglutinins, and those related to the carbohydrates and the lipids of the blood stream. Many of these globulins, each with a specific function, are present in such small amounts in the whole plasma as not to be readily detected. Their greatest value may well be as concentrates for the treatment of conditions in which there are deficiencies in these specific proteins.

Certain globulins are not very stable in solution. Accordingly, methods had been developed before the war for drying human plasma from the frozen state (18).[1] These have since been greatly simplified, and the safety and value of dried human plasma, one of the two blood substitutes thus far recommended by the National Research Council to the Armed Forces, has been amply demonstrated in all theatres of operation.

Plasma contains approximately 7 percent protein[2] and nearly 60 percent of this protein is albumin. Albumin is responsible for about 80

* The development of plasma fractions and the characterization of blood substitutes have been carried out in the Department of Physical Chemistry, Harvard Medical School, under contract recommended by the Committee on Medical Research between the Office of Scientific Research and Development and Harvard University since August 1941. Earlier work was supported by grants from the Rockefeller Foundation and in April 1941 grants from the Committee on Medicine of the National Research Council including a grant from the American College of Physicians were made available for the construction of the Harvard Pilot Plant for fractionation of plasma from blood collected by the American Red Cross.

The work of our department has been immeasurably aided by many colleagues—chemical, clinical, immunological and pathological—members of this or other Universities, of the Armed Forces, of Committees of the National Research Council, or of commercial laboratories with Navy contracts for plasma fractionation, who labored with us. Indeed, it would be impossible to list here the very large number of individuals and agencies who have collaborated in this work.

[1] This and similar notations refer to the articles in the References at the end of the paper.

[2] Blood is collected by the American Red Cross into a bottle containing a sodium citrate solution, and the citrated plasma thus contains about 6 percent protein.

percent of the colloid osmotic pressure (42b) which regulates the equilibrium in water and electrolytes between the plasma and the tissues, and thus plays the major role in maintaining the volume of the blood upon which normal circulation depends. More stable than plasma, albumin can be dispensed in solution (8c) (42a). Separated from the other plasma proteins by a new method developed in this laboratory (8b) (8g) (8h), albumin is now in full production in commercial laboratories under contract with the Navy.

Albumin is the least asymmetrical of the plasma proteins, the length of its molecule being only about four times its diameter. The globulin molecules are approximately twice as long as those of albumin, and fibrinogen molecules are six times as long, but all have about the same diameter. The more asymmetric the molecules of a substance, the more viscous its solutions and the more work the heart must do in maintaining the circulation of the blood if such a solution is injected.

Albumin solutions are so much less viscous than those of the other plasma proteins that a 25 percent solution has the same viscosity as whole blood (8d). This solution can be injected without reconstitution or dilution and will draw fluid from the tissues into the circulation in amounts which correspond to those predicted from measurements of its colloid osmotic pressure (24b) (42b). Since it can be transported as a concentrate without diluent, the standard Army-Navy package occupies but one-sixth the space of the equivalent dried plasma package. The chief precaution in the use of such concentrated albumin solutions is that in cases with severe dehydration, the necessary additional fluids should be given by mouth or by any other available route[3] (17) (26b) (49).

Human serum albumin is the second blood substitute recommended to the Armed Forces.[4] Its qualities of compactness and light weight are of obvious value for transportation over great distances in small ships, landing barges, or planes. It was developed to provide a stable blood derivative, instantly available for rapid administration, which meets the needs of highly mobile military units (24a) (26b) (35) (49).

There are special advantages respectively in the use of whole blood, of plasma, or of concentrated albumin solutions. In extensive hemorrhage, the need for both red cells and plasma proteins will be acute; in severe burns, where continuous replacement of plasma proteins is necessary, isotonic plasma has obvious advantages; and in acute surgical shock or hypoproteinemic edema, the strongly hypertonic albumin solution should draw water into the blood stream with maximum rapidity and with a minimum increase in its viscosity.

The injection of whole blood, plasma, or albumin (9) (47) may be used to restore blood volume in the prevention and therapy of shock. Shock as observed in military medicine generally results from a rapid decrease in the volume of circulating blood due to the loss of blood and plasma proteins externally and into damaged tissues. It is most frequently treated, or prevented, by injection of plasma proteins. The injected proteins, (1) insofar as they cannot readily pass the kidney, increase the body's reservoir of plasma proteins, and (2) insofar as they do not readily transverse the capillary walls, increase the plasma volume by drawing water from the tissues into the blood stream.

DIMENSIONS OF THE PLASMA PROTEINS AND OF CERTAIN SUGGESTED BLOOD SUBSTITUTES[5]

All of the plasma proteins that have thus far been carefully studied, here or elsewhere, have as their smallest dimension 33 Angstrom units or more (table I). They differ in length, the longer

[3] The Army-Navy package of normal human serum albumin which was developed by Capt. Lloyd R. Newhouser, Medical Corps, United States Navy, and Lt. Col. Douglas Kendrick, Jr., Medical Corps, United States Army, bears the statement "Precaution: in the presence of dehydration, albumin must be given with or followed by additional fluids."

[4] The clinical appraisal of human serum albumin was aided by the late Dr. S. Weiss and by Drs. R. F. Loeb, A. Blalock, I. S. Ravdin, O. H. Wangensteen, O. T. Bailey, D. W. Richards, Jr., E. A. Stead, Jr., Lt. Col. D. B. Kendrick, Jr., Lt. Comdr. S. T. Gibson, Lt. Comdr. L. M. Woodruff and many others and integrated largely by Dr. C. A. Janeway of our group and Capt. L. R. Newhouser of the Navy.

[5] The physical chemical characterization of a protein is a complex undertaking and we have been guided in this aspect of our work by ultracentrifuge and diffusion measurements carried out under the direction of Dr. J. L. Oncley in our laboratory and Dr. J. W. Williams in Wisconsin; by osmotic pressure, viscosity and nephelometry measurements under the direction of Dr. G. Scatchard at the Massachusetts Institute of Technology; by double refraction of flow measurements in the case of asymmetric molecules under the direction of Dr. J. T. Edsall; by rigidity of gels under the direction of Dr. J. D. Ferry; and by electrophoretic analyses carried out under the direction of Dr. S. H. Armstrong, Jr., in our laboratory and our collaborators in other laboratories. Our fractionations have been guided by our physical chemical characterization.

molecules which have the greatest effect on the viscosity of solutions being present in smallest amount; fibrinogen to the extent of 0.3 percent of the plasma. The equatorial diameters of the more symmetrical molecules, present in largest amount, would appear to be slightly greater, but as a first approximation all plasma proteins may be considered to have the same diameter, and this is the dimension which appears to prevent their rapid loss from the blood stream and is thus responsible for their sustained effect on blood volume.

Many of the physical and physical chemical properties of a protein of animal blood are very similar to those of the comparable protein of human blood. There are slight differences in chemical fine structure revealed by amino acid analyses (7) upon which observed differences in solubility and in antigenicity presumably depend. Albumins of bovine and human plasma are less antigenic than the corresponding globulins. Though bovine albumin, purified by crystallization (8f), has been as effective as human albumin upon injection and as free from immediate reactions (24b) (46), delayed reactions have occurred with certain preparations given to certain individuals. Therefore, at this time, we do not recommend for use in man even the best natural or modified products of animal blood that we have thus far investigated. Where sensitization to the foreign protein does not occur, however, the behavior of human and of bovine albumin in the body is remarkably similar as might be expected from their similarity of molecular dimensions. This appears to be true of metabolic availability as well as of retention in the blood stream, osmotic activity, and their consequent effectiveness in restoring and maintaining blood volume.

Hemodilution, an index of increase in the volume of circulating plasma, also follows the injection of concentrated salt or glucose solutions, but the osmotic effect of these small molecules is of short duration, since they readily leave the blood stream. The dimensions of the glucose molecule may be taken as 6.5 by 9.5 Angstrom units (table I).

Most so-called blood substitutes that have been suggested have diameters approximately one-half or less of those of the plasma proteins (table I) and leave the blood stream more readily than the plasma proteins. Glutamic acid peptide synthesized by *Bacillus subtilis* and separated as an interesting linear polymer (6) is estimated to be 11 Angstrom units in diameter. Gelatin, isinglass, and pectin are naturally occurring long, loosely coiled fibrous chains 15 to 18 Angstrom units in diameter.

Many observations (8d) suggest that molecules with a diameter greater than 30 Angstrom units do not under normal conditions pass into the urine, but that those with diameters of 20 Angstrom units or less leave the blood stream and are

TABLE I

ESTIMATES OF THE DIMENSIONS OF PLASMA PROTEINS AND OF SUGGESTED BLOOD SUBSTITUTES

	Molecular Weight	Dimensions of Ellipsoidal Model (Å = 10^{-8} cm.)	
		Length	Equatorial Diameter
Serum Albumin (Human)	69,000	150	38
Serum γ-Globulin (Human)	160,000	320	36
Fibrinogen (Human)	500,000	900	33
Serum Albumin (Bovine)	69,000	150	38
Globin (Human)	35,000	118	28
Gelatin	[a]	[a]	18
Isinglass	[a]	[a]	16
Glutamic Acid Peptide	15,000	260	11
Glucose	180	9.5	6.5
Chloride Ion	35.5	3.6	3.6
Sodium Ion	23	1.9	1.9

[a] Molecular weights and lengths vary with the degree of degradation of these materials, the maximum molecular weight and length being over 100,000 and 700 Å, respectively.

These estimates have been made by Drs. J. L. Oncley and A. Brown. Many of these values are derived from preliminary data. They are included because even approximate estimates of the size and shape of molecules are helpful in their consideration as "blood substitutes." The dimensions of the molecules are determined on the assumption that they are ellipsoids of revolution. The equatorial diameter is estimated from ultracentrifuge and viscosity measurements, in some cases supplemented by diffusion and double refraction of flow; the volume is estimated from the density and the osmotic pressure; the length of the molecule is estimated from the diameter and volume.

For heterogeneous materials such as glutamic acid peptide, autoclaved gelatin, isinglass, or pectin, the molecular weights and lengths are "number averages." Some of the molecules are very much lighter and shorter, but most of the weight is in molecules heavier and longer than this average. For a discussion of earlier literature and the constants on which certain of these calculations are made see (8c).

excreted in the urine, and the smaller their diameter, the more rapidly this occurs. Among molecules of the same diameter—though the electrical condition of the molecule and the membrane definitely influence permeability—elimination from the blood stream is generally more rapid the shorter their length.

Both gelatin and pectin are long rod-shaped molecules, in shape more like fibrinogen than like serum albumin or serum globulin (table I). Fibrinogen greatly increases the viscosity of solutions and causes pseudo-agglutination of red blood cells. Gelatin and pectin also have these effects. It was recognized that isosmotic solutions of both substances would be too viscous to be of value as blood substitutes. Both substances, however, are readily broken down into shorter rod-shaped particles. These degradation products appear to have approximately the same diameters but varying lengths. The problem in both cases was the same. Could a product be reproducibly obtained by carefully controlling the breakdown process in such a way that the particle size would become sufficiently small so that (1) viscosity would be reduced, and (2) the osmotic pressure increased without (3) so reducing the average particle size that its effect upon blood volume would not be sustained? In other words, when injected, would the product be a substitute for plasma proteins rather than for glucose and sodium chloride?

Solutions of gelatin that have been degraded would appear to be as satisfactory as any of the other suggested blood substitutes of this group. Gelatin properly prepared is non-antigenic and non-pyrogenic and has an effect in drawing water into the blood stream which is somewhat more sustained than that of salt or glucose solutions, though less sustained than that of the plasma proteins.[6] Degradation decreases the number of very long rod-shaped particles which are responsible for much of the viscosity and probably for other undesirable effects, while it increases the number of very small particles which are responsible for most of the osmotic pressure, though not for the colloid osmotic pressure, since they are rapidly lost from the blood stream. The longer rod-like particles appear to persist in the blood stream, and are presumably responsible for the increased sedimentation rate of the red cells, and possibly for changes in the distribution of plasma proteins between the blood stream and the tissues.

The albumins and globulins of plasma, though not spherical in shape, are generally termed globular in contrast to fibrous proteins. Many globular proteins, upon degradation, instead of yielding mixtures showing an almost continuous variation in size over a very wide range as do fibrous proteins, first dissociate to yield molecules of uniform dimensions, generally one-half, one-fourth, or one-sixth those of the original proteins (8e). This is true of hemoglobin, the respiratory protein of the red blood cell (table I). Since red blood cells contain approximately 30 percent hemoglobin and are a by-product of the Red Cross Blood Donor Service, large amounts of hemoglobulin are available and a great deal of physiological and some clinical work has been done with this natural product that has been suggested as a blood substitute (1a) (1b). From a chemical point of view it is homogeneous under most conditions with a molecular weight close to that of serum albumin. Its iron-containing, oxygen-combining pyrrole groups are readily transformed into a series of modifications, and the molecule itself is capable of division into two molecules of approximately equal size even without loss of its oxygen-combining capacity. Its elimination is far more rapid than would be expected of a protein the size of serum albumin and is reported to have been associated in certain cases with kidney damage. It remains a subject of controversy, however, whether the kidney damage was due to the proteins or the red cell residues.

Globin is derived from hemoglobin by extracting the iron-containing pyrrole groups in acid acetone (2). The protein precipitate, white in color, may be further treated in a variety of ways. Globin, like the products of the dissociation of hemoglobin into half molecules, with a molecular weight only half that of serum albumin, has a diameter (table I) more nearly comparable to the dimensions of the plasma proteins than are those of most other suggested blood substitutes.

THE EQUILIBRIUM BETWEEN THE ALBUMINS AND GLOBULINS OF PLASMA AND THE TISSUES

The plasma proteins are not merely molecules of a certain charge, shape, and size circulating within the blood vessels and maintaining the colloid osmotic pressure of the plasma and consequently the blood volume, but appear to be in

[6] Statement of the Subcommittee on Blood Substitutes of the National Research Council on the present status of studies on gelatin solutions prepared for intravenous use.

dynamic equilibrium with the tissues, from which they arise (26a) (29d) (32). The factors which regulate the concentrations and the total amounts of the different plasma proteins in the circulation are little understood, but with the preparation of plasma fractions containing individual protein components it should be possible to investigate these problems.

If plasma proteins are injected into a normal individual in excess of need, they leave the blood stream without appearing in the urine, and are presumably utilized as a source of energy or new protein or are stored in the tissues. The injection of concentrated serum albumin into the blood stream of a normal man increases the blood volume and serum albumin level only temporarily. Presumably an equilibrium between albumin circulating in the blood stream and the tissues leads to rapid redistribution of this normal plasma constituent with a return to a normal blood volume and a normal albumin level (26b). There thus appears to be an upper limit to the concentration and total amount of at least certain of the plasma proteins in normal blood. Whether this is regulated by colloid osmotic pressure, or more specifically for each species of plasma proteins, is less well known than it should be. Clearly there is an interaction between the various plasma proteins, which can be investigated by means of osmotic pressure measurements. The interaction in the body in health and disease can also be further investigated.

In pathological conditions, is albumin more likely than globulin to leave the blood stream? In the case of a small change in permeability in the surrounding membranes it surely is. Recent studies, both dielectric (19) and electrophoretic (28a) (28b) (28c) (29a) (29b) (29c) (31) (45), of the comparative distribution of the plasma proteins in blood and in urine or in body fluids suggest, however, that the old term, "albuminuria," was incorrect. Generally a number of different plasma proteins pass from the blood stream. The electrophoretic distribution of protein components appearing in the urine under certain pathological conditions is more nearly that of normal plasma than that of pathological plasma. This implies a greater loss of albumin from the plasma than of globulin, especially in the case of an albumin-poor plasma. In the case of serous effusions also, "albumin is present in the same or higher proportion, in the effusion, as compared with plasma, and beta globulin and fibrinogen in lower proportion" (29c). It would be important to know whether the deductions drawn from the study of serous effusions apply equally to the oozing from burns.

From a physical chemical point of view it is possible to understand these phenomena, at least as a first approximation. Since all of the plasma proteins have about the same equatorial diameter, any change in the permeability of a membrane which would permit one plasma protein to diffuse through it should permit any of them to do so. Under these circumstances the amount of plasma protein of each species to diffuse should be an inverse function of the length of its molecules (table I). Consequently for a given change in permeability there will be a greater loss—and therefore a greater need for replacement—of albumin than of the globulins or fibrinogen.

Clearly there is a principle guiding the amount of any plasma component that may wisely be used in replacement therapy, namely, the amount that will tend to re-establish rather than to derange the normal concentrations of the various physiologically significant plasma components.

THE DISSOCIATION AND AGGREGATION OF THE GLOBULIN MOLECULES OF PLASMA

The mass of the albumin molecule does not appear to alter under different conditions. The mass distribution of certain of the proteins of plasma or serum, however, has previously been noted to change with protein and electrolyte concentrations. There is an alpha globulin fraction (α_1) which appears to dissociate into molecules approximately the size of albumin and to resemble albumin also with respect to renal clearance (29d). There is also alpha globulin which aggregates into molecules larger than normal globulin. These components of alpha globulin, noted in a fraction of horse serum (8a), have also been noted in our fraction (IV-1) of human plasma. It is the alpha globulins which are increased in febrile conditions, and in lipoid nephrosis, in which fibrinogen increases in the plasma as do the alpha and beta globulins, whereas the albumins, and to a lesser degree the gamma globulins, are diminished. The importance of studying the variation in concentration of these proteins of variable molecular weight, both in the laboratory and in the clinic, can scarcely be overemphasized.

Whereas normal gamma globulins of mass less than twice that of albumin have not been noted, they have been prepared by enzymatic processes in the laboratory (30) (38a) (38b) and may well

have value in the treatment of certain infectious diseases. In hyper-immunized animals, and presumably in man, the gamma globulins aggregate however, so as to yield molecules of much larger molecular weight (23a) (23b) (27a) (27b). Their aggregation appears to be accomplished by end to end conjugation, since such gamma globulins reveal double refraction of flow (48b). Large increases in the amount of antibodies have been associated with large increases in the amount of globulin, and in immunized animals this has often proceeded so far as to produce an absolute increase in the total serum proteins.

These interesting results suggest how the process of immunization, resulting in molecules of large molecular weight and very small net charge at physiological reactions, can take place without serious disturbances of blood volume. These large globulins, existing close to their isoelectric points, make but relatively small contributions to the osmotic pressure of the plasma.

Fragmentary as the present state of knowledge regarding the interactions of the plasma proteins is, a number of mechanisms in terms of which equilibration may be effected have been noted. We have seen that increase in the immune gamma globulins is sometimes accompanied by increase in molecular size, thus minimizing the effect on colloid osmotic pressure. Alpha globulins of increased molecular weight have also been observed, as have alpha globulins more nearly the size of albumin, which resemble albumin not only in this respect but also in renal clearance. Albumin, though it does not change in size, rapidly redistributes itself between plasma and tissues, and this may be true of certain other plasma proteins.

What is the effect on these complex equilibria of the injection of large amounts of plasma proteins? And what will be the effect of the injection of the various suggested blood substitutes? Can colloids force certain of the plasma proteins from the blood stream, thus bringing about a redistribution of the plasma proteins? There is evidence that the injection of gum acacia has resulted in a rapid fall of serum albumin concentration (22), suggesting that foreign colloids not readily utilized in the body may displace the normal colloids of the plasma. How general are these phenomena? These questions are raised by the increasing use of transfusions in therapy and can now be far more specifically investigated than has heretofore been possible.

THE SEPARATION AND CONCENTRATION INTO FRACTIONS OF THE ALBUMINS AND GLOBULINS OF PLASMA[7]

In order to make available each of the protein components of plasma for therapeutic uses based on knowledge of their physiological functions, it was necessary to develop a large scale fractionation process. The potential value of all of the different proteins of human plasma demanded that this process be inclusive in the sense that all components be separated and concentrated. The present process, by which the blood of more than a million donors has been treated, separates the plasma into five fractions. Fraction I is rich in fibrinogen. Fraction II + III contains essentially all of the immune globulins as well as prothrombin and the isohemagglutinins. Fraction IV contains most of the alpha globulins. Fraction V contains between 85 and 90 percent of the albumin. The mother liquors in Fraction VI contain but 1 percent of the total proteins. The yields of the proteins in the various fractions are given in table II.

Fibrinogen and the gamma globulins are completely distributed between Fraction I and II + III. Since there have in the past been more obvious uses of immune globulins than of fibrinogen, the separation between these two fractions has thus far been carried out so as to lose no more gamma globulin into Fraction I than is represented by occlusion, losing rather fibrinogen into Fraction II + III. Slight changes in the process are under investigation to increase the amount of fibrinogen in Fraction I, and its purity may be readily increased by reprecipitation.

Prothrombin is concentrated in Fraction II + III. Its functions are discussed elsewhere in this paper. A subfractionation of II + III concentrates the prothrombin in a small fraction, III-2, essentially free of antibodies. This fraction, which also contains the mid-piece of complement, C′1[8] is the source material for further purification of prothrombin and thus of thrombin.

Fraction II + III contains approximately equal parts of beta and gamma globulin. A

[7] The continuous operation of our pilot plant for two and a half years has made possible the preparation of each product developed and its standardization for clinical trial. The work of the pilot plant has been carried out under the direction of Dr. L. E. Strong with the co-operation of Dr. W. L. Hughes, Jr., Dr. D. J. Mulford, Mr. M. Melin, Mr. J. N. Ashworth, Mr. J. H. Cameron, Mr. H. T. Gordon, Lt. H. L. Taylor, Dr. D. A. Richert, Mr. R. W. Kelty, Mr. L. H. Larsen.

[8] Studied in collaboration with Lt. Louis Pillemer of the Army Medical School.

TABLE II[a]

DISTRIBUTION OF PROTEINS OF PLASMA[b]

	Grams Protein per Liter Plasma	Grams Protein per Liter of Plasma in Fractions					Protein in Fractions
		I	II + III	IV	V	VI	
Protein	60.3[c]	4.3	16.3	9.7	29.6	0.6	60.5
Albumin	33.2	0.2	0.7	1.0	29.0	0.3	31.2
α-globulin	8.4	0.2	1.8	5.4	0.6	0.3	8.4
β-globulin	7.8	0.8	6.2	3.1			10.1
γ-globulin	6.6	0.5	6.0	0.2			6.7
Fibrinogen	4.3	2.6	1.6[d]				4.2
Totals	60.3	4.3	16.3	9.7	29.6	0.6	60.6

[a] Reprinted from *Jour. Clin. Invest.* **23**: 417, July 1944, with permission of the editors.

[b] The distribution of components in plasma and each fraction is based on electrophoretic analysis which gives the fraction of the total refractive increment contributed by each component. The size of each fraction is based on nitrogen analysis assuming a nitrogen factor of 6.25. It has been assumed that the refractive increment per gram of nitrogen is the same for all components. Work is in progress to determine the relations among nitrogen content, refractive increment, and dry weights for the various proteins.

[c] The nitrogen values have been deduced by Dr. H. B. Vickery, who has been in charge of the calculation of yields in plasma fractionation as carried out at Harvard and the commercial firms with Navy contracts, from the study of 34 large pools of citrated plasma. For the total nitrogen in plasma the average value is 9.88 g. per liter. Of this the non-protein nitrogen is 0.24 g. per liter leaving 9.64 g. of protein nitrogen per liter. Using 6.25 as the nitrogen factor gives 60.3 g. of protein per liter.

[d] The electrophoretic analysis of Fraction II + III has been corrected to include 10 percent of fibrinogen precipitated as fibrin previous to analysis.

large part of the beta globulin which is isoelectric near pH 5.2 (13a) (13b) (20) (33a) (33b) (43) is separated into Fraction III-2 which generally contains less than 15 percent gamma globulin and more than 75 percent of beta globulin. Indeed, preparations have been made which are more than 90 percent beta globulin.[9]

The isohemagglutinins are also found in Fraction II + III[10] and may be concentrated into a subfraction III-1. When prepared from typed blood, either Fraction II + III or III-1 may therefore serve as a source for the further purification of isohemagglutinins (40c) of value in the typing of whole blood.

The beta globulins, present to approximately 60 percent in Fraction III-1, are largely isoelectric near pH 6.3, and some of them appear to be associated with a portion of the lipids of the plasma.

The immune globulins[11] have been concentrated in Fraction II + III. The concentration of antibodies in Fraction II + III should have been increased ten-fold as compared to the plasma provided there was no destruction of activity during the process of separation and concentration. The results in table III demonstrate—within the limits of error of immunological technique—that this has been accomplished (table III).

Fraction III-1 is not free of antibodies. Indeed the anti-typhoid O agglutinin is concentrated in Fraction III-1 which is a mixture of beta and gamma globulins. The gamma globulins represent approximately 25 percent of the crude fraction and account for many at least of the antibodies present.

The gamma globulins. Essentially all of the antibodies in the original plasma, and in Fraction II + III are concentrated in Fraction III-1 or Fraction II. Fraction II may be prepared substantially free of cholesterol and of beta globulin, prothrombin, thrombin, and isohemagglutinins. Its content of gamma globulins is generally over 90 percent, but by changing the antibody distribution between Fractions II and III-1 preparations containing 99 percent or more gamma globulins are readily prepared. Clinical studies of such preparations have shown that they contain effective concentrations of protective antibodies against measles (26d) (41). Such preparations also have titers of certain other antibodies at least as high, and presumably higher than those poorer in gamma globulin. It would thus appear that many of the antibodies in human blood are gamma globulins and that Fraction II is the purest source of such antibodies now available.

The alpha globulins are concentrated in Frac-

[9] These studies on thrombin and prothrombin have been largely carried out by Dr. J. T. Edsall and Miss Susan Miller.

[10] The separation of isohemagglutinins with this fraction was first noted by Dr. W. C. Boyd in 1941 and the further purification developed in collaboration with the Army Medical School.

[11] We are indebted to Dr. A. R. Dochez for making arrangements for the initial immunological survey of our fractions which have been intensively investigated by Dr. J. F. Enders (15).

TABLE III [a]

ANTIBODIES IN GLOBULIN FRACTION II + III DERIVED FROM HUMAN PLASMA

Antibody	Type of Antibody	Investigator	Institution	Concentration ref. to plasma
Anti-diphtheria	Antitoxin neutralizing	Edsall	Mass. Antitoxin and Vaccine Laboratory	10
Anti-dysentery	Agglutinins	Mudd	Univ. of Pennsylvania	2–10
Anti-herpes simplex	Neutralizing	Stokes	Children's Hosp., Phila.	[b]
Anti-influenza (human PR8)	Hirst inhibition	Hirst	Rockefeller Institute	4–8
Anti-influenza (human PR8)	Hirst inhibition	Eaton	Calif. Dept. Public Health	4
Anti-influenza (human PR8)	Hirst inhibition	Enders	Harvard Med. School	10–15
Anti-influenza (human PR8)	Complement fixation	Enders	Harvard Med. School	10–15
Anti-influenza (human PR8)	Neutralizing	Stokes	Children's Hosp., Phila.	10
Anti-influenza (human PR8)	Neutralizing	Enders	Harvard Med. School	9
Anti-influenza (swine)	Neutralizing	Stokes	Children's Hosp., Phila.	10
Anti-influenza (swine)	Neutralizing	Shaffer	Mass. Antitoxin and Vaccine Laboratory	4
Anti-lymphocytic choriomeningitis	Neutralizing	Stokes	Children's Hosp., Phila.	[b]
Anti-measles	Protective (human)	Stokes	Children's Hosp., Phila.	[b]
Anti-mumps	Complement fixation	Enders	Harvard Med. School	2–10
Anti-parapertussis	Agglutinins	Mudd	Univ. of Pennsylvania	64
Anti-pertussis	Agglutinins	Mudd	Univ. of Pennsylvania	4–10
Anti-pertussis	Agglutination	Enders	Harvard Med. School	10
Anti-pertussis	Mouse protection	Bradford	Univ. of Rochester	4–10
Anti-perfringens	Protective	Hall	Univ. of Colorado	[b]
Anti-poliomyelitis	Neutralizing	Kramer	Mich. Dept. Health	10
Anti-poliomyelitis	Neutralizing	Stokes	Children's Hosp., Phila.	16
Anti-poliomyelitis	Rat and mice protection	Kramer	Mich. Dept. Health	10
Anti-poliomyelitis	Rat and mice protection	Stokes	Children's Hosp., Phila.	10
Anti-scarlatina	Neutralizing	Bradford	Univ. of Rochester	[b]
Anti-scarlatina	Neutralizing	Wadsworth	New York Dept. Health	5–10
Anti-streptococcus	Antitoxin	Wadsworth	New York Dept. Health	4–10
Anti-typhoid	H agglutinin	Enders	Harvard Med. School	8–10
Anti-typhoid	O agglutinin	Enders	Harvard Med. School	8–10
Anti-vaccinia	Neutralizing	Janeway	Children's Hosp., Boston	[b]
Isoagglutinins	Agglutinins	Boyd	Harvard Med. School	8–10

[a] Reprinted from *Jour. Clin. Invest.* **23**: 510, July 1944, with permission of the editors.
[b] Activity present but no quantitative data.

tion IV. Alpha globulins of two kinds, α_1 and α_2, are recognized electrophoretically, and subfractionation yields preparations (IV-1) consisting of approximately 90 percent α_1.

Fraction IV contains the end-piece of complement (40a) (40b), hypertensinogen, thyrotropin and presumably other hormones and enzymes. A smaller fraction than II + III, its components are less well known, both with respect to their chemical nature and their physiological functions. The separate fractions of alpha globulin are now beginning to be studied however and their nature and value appraised.

Albumin is concentrated in Fraction V which is free of fibrinogen, gamma globulins and all but traces of beta globulin and cholesterol. According to the specifications in the Navy contracts for the preparation of this blood derivative albumin may contain up to 2 percent of globulin. On the average the albumin preparations accepted have contained 1.5 percent alpha globulin (48a).[12] These small amounts of globulin can be removed by crystallization of the albumin. Albumin of this purity is readily prepared and has been used therapeutically. Its stability is higher but its preparation involves additional processes and there appears to be no clinical advantage in this further purification.

[12] It is not impossible that in the future inclusion of larger amounts of alpha globulins separated in Fraction IV may be permitted, provided their stability and safety have been demonstrated, and provided more specific uses for these components of plasma are not discovered.

THE STABILITY OF THE ALBUMINS AND GLOBULINS SEPARATED FROM PLASMA

Plasma is by no means as stable as the proteins that can be separated from it. Indeed environments are being found in which certain of the proteins separated from the plasma are more stable than they are in each other's presence.

Albumins are the most stable of the proteins thus far investigated. It is for this reason that albumins can be dispensed in solution. Ever since these investigations were begun, advance had followed advance in determining conditions such that maximum stability would be achieved. First the pH of the 25 percent solution was adjusted to 6.8 (42a). Next it was discovered that thermal stability increased almost linearly with salt concentration, and the sodium chloride concentration in the standard package was increased to 0.3 molal (42a). Next the influence of nonpolar anions on the thermal stability of serum albumin was demonstrated by Luck and his coworkers (3a) (3b). The replacement of approximately 0.04 or 0.05 M. of the chloride ion concentration with phenyl acetate or mandelate anion was found by them to increase from five to eight fold the length of time during which albumin can be subjected to high temperatures without change.* That this stabilizing effect is on the albumins and not on the globulins can be shown in the case of bovine albumin by the demonstration, using a sensitive immunological technique, that such solutions have become free of globulin after heat treatment and filtration. These observations thus render possible new developments in the purification of both human and animal albumins (table IV).

TABLE IV

INFLUENCE OF SOLVENT COMPOSITION UPON THE RELATIVE THERMAL STABILITY OF CERTAIN PLASMA PROTEINS AT 57° C.

Composition of Solvent (moles non-protein solute per liter of solution)	Human Serum Albumin[a] 25% solution	Human Serum γ-Globulin[b] 20% solution
0.6 M Glycine	0.1	7
0.3 M Sucrose	—	7
0.15 M Glycine + 0.15 M Sucrose	0.6	5
0.3 M Glycine	—	4
Water	0.1	1.0
0.15 M NaCl	0.5	1.0
0.3 M NaCl (Taken as unity)	1.0	1.0
0.15 M NaButyrate + 0.15 M NaCl	>10	0.8
0.15 M NaMandelate + 0.15 M NaCl	>10	0.5
0.15 M NaPhenylacetate + 0.15 M NaCl	>10	0.4

[a] Relative stabilities estimated from measurements of Drs. L. E. Strong, W. L. Hughes, Jr., and A. Sparrow, Harvard Medical School. These results confirm those of Dr. J. M. Luck and his co-workers, Stanford University (3a) (3b).

[b] Relative stabilities estimated from measurements of Mr. P. Gross, Jr., Department of Physical Chemistry, Harvard Medical School, and of Dr. A. Brown, Department of Physical Chemistry, Massachusetts Institute of Technology.

The physical and chemical properties of the albumins and gamma globulins cannot be more dramatically contrasted than in terms of the environments in which their thermal stability is greatest. In contrast to the albumins, the thermal stability of gamma globulin is almost independent of salt concentration. Salts of nonpolar anions, instead of increasing, decrease, whereas strong polar groups, such as those borne by glycine and sugars, increase thermal stability (table IV).[13] In this connection it should be noted that gamma globulins have the greatest electric moments of any of the plasma proteins thus far investigated (16a) (19) (36) as well as the highest concentration of free hydroxyl groups (7) (one in every four of the residues being an hydroxy-amino acid according to Brand's analyses). These investigations, which practical considerations demanded, thus appear to suggest a general principle which if demonstrated could be of far-reaching theoretical importance; namely that a molecule is most stable when surrounded by other molecules possessing many of the same groups or properties. That like dissolves like is an old axiom; the hypothesis which we are now adopting is that like stabilizes like and its implications extend beyond physical chemistry into physiology and morphology.

THE USE OF THE GAMMA GLOBULIN ANTIBODIES IN THE PREVENTION AND MODIFICATION OF MEASLES

The antibodies of a pooled plasma reflect the degree of immunity to various infectious diseases of the community contributing the blood. This might thus be expected to vary from continent to

* Note added to proof: Most recently the use of acetyl phenylalanine, suggested by Dr. H. Clarke, has been demonstrated to be effective and Dr. L. E. Strong has suggested the use of comparable derivatives of those amino acids in which the protein should be supplemented for growth or plasma regeneration.

[13] These observations have been made by Mr. Paul Gross in our laboratory.

continent and from time to time. The plasmas that have been processed and studied by us during the past years have been derived from blood collected in various parts of the United States and have shown fairly constant distribution of antibodies to the more common diseases for which tests have been made (15).

The possibility of utilizing the gamma globulins derived from pooled normal plasma in the prophylaxis and therapy of any infectious disease depends upon two factors: first, the concentration of specific antibody in the blood of the adult population, which in turn is dependent on previous exposure and on the persistence of these bearers of immunity in the body and their distribution between blood and tissues; second, the amount of antibody necessary to protect an individual against the infection by passive immunization. The concentration over the pooled plasma thus far achieved by our fractionation process is approximately twenty to thirty fold if we compare plasma with the concentrated solution of gamma globulins being made available to the Armed Forces.* As a result titers of certain antibodies, though not necessarily of all, are comparable to those of the corresponding convalescent sera. Therefore, in the prevention or treatment of any disease where the value of convalescent serum has been demonstrated, the effectiveness of a concentrate from pooled normal plasma should be investigated. If convalescent serum, or hyper-immune serum is fractionated, still higher antibody titers may be obtained in the concentrated gamma globulins, and their possible value in special cases must also be considered.

TABLE V

MEASLES PROPHYLAXIS

Comparative Results with Different Agents Used for Passive Protection. (No selection according to age, dose, or exposure-injection interval)

	Number of Cases	Success* %	Failure** %
Normal Adult Serum[a]	584	80	20
Convalescent Measles Serum[a]	1627	92	8
Placental Globulin[a]	2740	95	5
Immune Serum Globulin	1687***	97	3

* Combined figures for those protected and those having mild measles.

** Figures for those with average measles.

*** The following chart is a breakdown of this total.

Season	Group	No.	No Measles	Mild Measles	Average Measles
1942–43	Stokes, Maris, and Gellis (44)	891	618	248	25
1942–43	Ordmann, Jennings, and Janeway (37)	281	166	104	11
1943–44	Rutstein, Greenberg, and Frant[b]	404	335	65	4
1943–44	Stokes, Gellis, and Janeway, et al.[c]	111	46	57	8

[a] McKhann, C. F. The prevention and modification of measles. *Jour. Amer. Med. Assoc.* **109**: 2034–2037, 1937.

[b] Rutstein, D. D., Greenberg, M., and Frant, S. Work carried on by the Department of Health, New York City, and reported by them to us.

[c] Work carried on by local physicians in Boston, Philadelphia, and Baltimore, and reported to us by Drs. J. Stokes, Jr., and E. P. Maris, Lt. S. S. Gellis, and Dr. C. A. Janeway. We are indebted to Dr. C. A. Janeway for integrating these appraisals and preparing this table which summarizes the results in (37) (44) as well as those observed more recently.

The gamma globulins concentrated from pooled plasma have proven very effective in the prevention and modification of measles. The results of studies on their use in this disease are summarized in table V (37) (41) (44).

THE USE OF THE ISOHEMAGGLUTININS IN THE TYPING OF BLOOD

The isohemagglutinins represent but a small part of Fraction III-1 in which they separate in our process. They are euglobulins, isoelectric near pH 6.3 and far less sensitive to ionic strength at more acid than at more alkaline reactions.[14] Further investigation of their solubility relations should permit of their further purification and stabilization.

The development of the isohemagglutinins of pooled plasma for typing of whole blood, carried out at the suggestion of and with the collaboration of the Army Medical School, has reached a point where production for the Armed Forces has begun (40c). In the effort to standardize these products of plasma fractionation, reference standards have been prepared at the Harvard Pilot Plant and made available to those concerned with the appraisal (12) of these isohemagglutinins. Their results have once more

* Note added to proof: Gamma globulin antibodies, by-products of serum albumin prepared under Navy contract, will be made available to the civilian population by the American Red Cross through public health agencies insofar as they are in excess of military needs (*Jour. Amer. Med. Assoc.* **125**: 635, July 1944).

[14] These studies are being carried forward by Mr. M. Melin and Mr. J. Cameron in our laboratory.

brought out the extent to which the phenomenon of red cell agglutination depends not only upon variations in technique but also upon variations in the properties of red cells even of the same type (12).

THE USE OF FIBRINOGEN AND THROMBIN IN THE PRODUCTION OF CLOTS OF DIVERSE PHYSICAL PROPERTIES

Fibrinogen and thrombin, concentrated in the course of the plasma fractionation process into Fractions I and III-2, respectively, react to form a solid structure, the fibrin clot[15] (14). By controlling the conditions of the reaction, structures of a wide variety of properties can be obtained (16b). Clots of one type are clear, gelatin-like, elastic, and friable, do not readily synerize with loss of fluid, and adhere to surfaces on which they are formed. Clots of another type are opaque, doughy, and scuff-resistant, synerize readily, become dense with loss of fluid, and do not adhere to underlying surfaces. Clots of intermediate properties can also be made.

For surgical use in human patients, clots may be prepared from fibrinogen and thrombin to meet the particular specifications chosen for each type of application, and may be formed *in situ* in or over wounds. Clots the properties of which have been controlled in this way have been used in pyelolithotomy (11) and in a technique for skin grafting (10).

USE OF FIBRIN FILMS AS DURAL SUBSTITUTES AND IN OTHER SURGICAL APPLICATIONS

The reaction between fibrinogen and thrombin *in vitro* can be employed to prepare still other products, which bear less resemblance to normal clots, but have found uses in surgery and medicine. Just as the asymmetry of certain synthetic polymers is responsible for the remarkable mechanical properties of industrial plastics, the fibrinogen molecule endows these products of human proteins with remarkable mechanical properties.

Fibrin film has been prepared from fibrinogen and thrombin as a strong, rubbery sheet which can be stretched reversibly from two to three times its original length (16b). This material can be made in various shapes and thicknesses, and in the form of seamless tubing. Its mechanical properties can also be varied from a soft, rubberlike elasticity to a parchment-like consistency.

The mechanical properties of the soft, rubbery films (as shown by their stress-strain curves) bear a marked similarity to those of the elastic ligament of the neck, the *ligamentum nuchae*. The tough films with delayed elasticity resemble rather wool or hair in their mechanical properties. It may be possible to imitate, with different types of fibrin products, still other natural structures of the body (16b).

The fine structure of fibrin film has been shown to involve pores which in one type are of the order of 30 Angstrom units or more in diameter. Hemoglobin molecules in solution pass through these pores readily, but plasma globulin molecules are partially, and fibrinogen molecules largely, retained.

The mechanical properties of certain types of fibrin films make them suited for use as dural substitutes and in the prevention of meningocerebral adhesions. The duration of their persistence in the body can be adjusted by suitable treatments (34). They have been used in 52 neurosurgical operations by our group as well as by others. Experience has indicated that fibrin film is an excellent material for these purposes (4b) (25a) (25c), patients having been followed for as long as ten months without the appearance of unfavorable sequelae.

Other types of film are being tested for other surgical applications, including the treatment of burns (21).

THE USE OF FIBRIN FOAM AND THROMBIN AS A HEMOSTATIC AGENT

For stopping the flow of blood, in surgical procedures, thrombin is the only component of the clotting mechanism which must be supplied. For the most effective use of thrombin in hemostasis, however, it must be applied with a matrix which can prevent the blood flow from washing the thrombin away before the clotting is completed. A porous matrix soaked in thrombin solution can accomplish this mechanical result (39). Fibrin foam is such a matrix formed from human fibrinogen and thrombin (5a).

In the dry state fibrin foam is a porous material composed of strands of fibers separated by air spaces of macroscopic size. Its density is of the order of 0.03 g. per cc. It readily absorbs water, saline, or thrombin solution. The physical

[15] In certain pathological conditions such as hemophilia and purpura hemorrhagica, clots of abnormal properties and consistencies have been described. (See Edsall, G., and McKhann, C. F., *Penna. Med. Jour.* **42**: 731, April 1939).

properties of the foam can be varied by controlling the conditions of manufacture. One type is light, fluffy, and highly compressible; it wets with ease and, when wet, loses about 90 percent of its dry volume through spontaneous shrinkage. Another type is dense, firm, and less compressible; it wets slowly, and, when wet, shrinks only about 50 percent of its dry volume.

The material effectively controls bleeding from oozing surfaces and from veins, even very large ones. It is not recommended for brisk arterial hemorrhage. Clinical experience thus far has been greater with its use in neurosurgery than elsewhere, fibrin foam having been used as a hemostatic agent in 143 neurosurgical operations by our group and in many more in other clinics.[16] It has proved very effective in controlling hemorrhage from the dura, from tumor beds, from dural sinuses, and from large cerebral veins (4a). Since fibrin foam is rapidly absorbed with minimal tissue reaction, it is left in place, thus preventing recurrence of the bleeding. In comparison with muscle, fibrin foam causes much less tissue reaction,[17] is easier to handle in the operating room, and is immediately available for use in emergencies (25b).

Satisfactory use of fibrin foam has been made in general surgery, but thus far experience is limited to a small number of patients. It has proved effective in controlling hemorrhage from the cut surface of the liver and kidney, in jaundiced patients, and in thoracic operations. Fibrin foam has also been of great value in stopping bleeding in hemophilia when other methods have failed (5b) (25c).

SUMMARY

Blood substitutes of many kinds have been suggested. The synthetic polymers and natural chain-like molecules that have been investigated have diameters of 18 Angstrom units or less, as compared with diameters of 33 Angstrom units or more for the plasma proteins. Although all have the property of drawing water into the blood stream, the rate at which they are lost from the blood stream varies greatly, and appears to depend (1) upon the diameter of the molecules and (2) upon their length among molecules of the same diameter. The longer the molecule, however, the greater the viscosity of its solutions and the more work is imposed upon the heart. Most long asymmetric molecules that have been studied also have a tendency to bring about clumping of red blood cells. The influence of foreign colloids introduced into the blood stream upon the plasma proteins should also be considered.

Human blood and blood derivatives are playing a significant role in military medicine in this war. No blood substitute that has been considered completely replaces whole blood. The red blood cells which carry oxygen to the tissues are especially needed when there has been great blood depletion, when major operations are to be performed, or when anemia must be combatted during convalescence.

Plasma, readily separated from the more labile cellular constituents of the blood, is rendered stable by drying. In this form large amounts have been prepared from the blood collected by the American Red Cross and accumulated against military needs.

Plasma is further separated into a series of protein products, each a stable dry white powder responsible for a specific natural function. Each product can be used as a specific therapeutic agent. Such fractionation of the blood collected by the American Red Cross, is also being carried out on a large scale. The products of human plasma thus far proven of value in medicine and surgery include:

Albumin which represents less than 60 percent of the plasma proteins but is responsible for almost 80 percent of the osmotic activity of plasma which maintains the equilibrium in water and electrolytes between blood and the tissues. The injection of albumin or plasma is used in the prevention or treatment of shock, since these large protein molecules, which do not normally pass the kidneys, draw water into the blood stream and thus restore blood volume and the normal circulation of the blood.

Gamma globulins, the concentrated antibodies of human blood, have been separated in a small fraction and made available in a solution concentrated twenty-five fold with respect to the plasma. They have proven of value in the prevention and modification of measles.

Isohemagglutinins, separated from plasma frac-

[16] The clinical appraisal of fibrin foam and thrombin has been greatly aided by Drs. H. C. Naffziger, G. Horrax, C. Pilcher, and C. Coleman of the Subcommittee on Neurosurgery of the National Research Council, as well as by Drs. J. Evans and E. Sachs, Lt. Col. R. G. Spurling, Medical Corps, United States Army, and Captain W. M. Craig, Medical Corps, United States Navy, and has been integrated by Dr. F. D. Ingraham of our group.

[17] Histological and pathological studies of all our products have been carried out under the direction of Dr. O. T. Bailey.

tionated from group specific blood, have proven of value in blood grouping.

Fibrinogen, the long rod-shaped molecules which represent only 6 percent of the plasma proteins, upon which the physical properties of the blood clot depend, can be made into plastics or in connection with *thrombin* into clots with a variety of physical properties of value in a variety of conditions.

Fibrin films prepared from the clots are being used as dural substitutes in brain operations and studied as replacements for other naturally occurring membranes.

Fibrin foams, constituted of the same human proteins as the natural blood clot, are being used with thrombin as hemostatic agents.

The separation and concentration of the various proteins of human plasma continues in order to render as many as possible available as therapeutic agents, and thus to increase our knowledge and control of the composition of the blood in health and disease.

REFERENCES

1. a. Amberson, W. R. 1937. Blood substitutes. *Biol. Rev.* **12**: 48–86.
   b. Amberson, W. R., Jacobs, J. E., Hisey, A., and Monke, J. V. 1942. Hemoglobin-saline solutions as transfusion media. Chapter XIX in *Blood Substitutes and Blood Transfusions* by Mudd, S., and Thalhimer, W. Charles C. Thomas, Baltimore.
2. Anson, M. L., and Mirsky, A. E. 1931. Protein coagulation and its reversal. Globin. *Jour. Gen. Phys.* **14**: 605–609 (May).
3. a. Ballou, G. A., Boyer, P. D., Luck, J. M., and Lum, F. G. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. V. The influence of non-polar anions on the thermal stability of serum albumin. *Jour. Clin. Investigation* **23**: 454–457 (July).
   b. Ballou, G. A., Boyer, P. D., Luck, J. M., Lum, F. L. 1944. The heat coagulation of human serum albumin. *Jour. Biol. Chem.* **153**: 589–605 (May).
4. a. Bailey, O. T., and Ingraham, F. D. 1944. Chemical, clinical, and immunological studies on the products of human plasma fractionation. XXI. The use of fibrin foam as a hemostatic agent in neurosurgery: Clinical and pathological studies. *Jour. Clin. Investigation* **23**: 591–596 (July).
   b. Bailey, O. T., and Ingraham, F. D. 1944. Chemical, clinical, and immunological studies on the products of human plasma fractionation. XXII. Fibrin films in neurosurgery, with special reference to their use in the repair of dural defects and in the prevention of meningocerebral adhesions. *Jour. Clin. Investigation* **23**: 597–600 (July).
5. a. Bering, E. A., Jr. 1944. Chemical, clinical, and immunological studies on the products of human plasma fractionation. XX. The development of fibrin foam as a hemostatic agent and for use in conjunction with human thrombin. *Jour. Clin. Investigation* **23**: 586–590 (July).
   b. Bering, E. A., Jr. 1944. Clinical uses of products made from human fibrinogen and human thrombin. *Bull. U. S. Army Med. Dept.* **78**: 53–56 (July).
6. Bovarnick, M. 1942. Formation of extracellular d(-)-glutamic acid polypeptide by bacillus subtilis. *Jour. Biol. Chem.* **145**: 415–425 (Oct.).
7. Brand, E. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. III. Amino acid composition of plasma proteins. *Jour. Clin. Investigation* **23**: 437–444 (July).
8. a. Cohn, E. J., McMeekin, T. L., Oncley, J. L., Newell, J. M., and Hughes, W. L., Jr. 1940. Preparation and properties of serum and plasma proteins. I. Size and shape of proteins separating upon equilibration across membranes with ammonium sulfate solutions of controlled pH, ionic strength and temperature. *Jour. Amer. Chem. Soc.* **62**: 3386–3393 (Dec.).
   b. Cohn, E. J., Luetscher, J. A., Jr., Oncley, J. L., Armstrong, S. H., Jr., and Davis, B. D. 1940. Preparation and properties of serum and plasma proteins. III. Size and charge of proteins separating upon equilibration across membranes with ethanol-water mixtures of controlled pH, ionic strength and temperature. *Jour. Amer. Chem. Soc.* **62**: 3396–3400 (Dec.).
   c. Cohn, E. J. 1941. The properties and functions of the plasma proteins with a consideration of the methods for their separation and purification. *Chem. Rev.* **28**: 395–417 (April).
   d. Cohn, E. J. 1942. The plasma proteins: their properties and functions. *Trans. & Studies Coll. of Phys. of Philadelphia*, 4 Ser. **10**: 149–162 (Dec.).
   e. Cohn, E. J. 1939. Some physical-chemical characteristics of protein molecules. *Chem. Rev.* **24**: 203–232 (April).
   f. Cohn, E. J., and Hughes, W. L., Jr. Unpublished data.
   g. Cohn, E. J., Oncley, J. L., Strong, L. E., Armstrong, S. H., Jr., Ferry, R. M., and Hughes, W. L., Jr. 1942. Properties and functions of the purified proteins of animal and human plasmas. Chapter XX in *Blood Substitutes and Blood Transfusions* by Mudd, S. and Thalhimer, W. Charles C. Thomas, Baltimore.
   h. Cohn, E. J., Oncley, J. L., Strong, L. E., Hughes, W. L., Jr., and Armstrong, S. H., Jr. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. I. The characterization of the protein fractions of human plasma. *Jour. Clin. Investigation* **23**: 417–432 (July).
9. Cournand, A., Noble, R. P., Breed, E. S., Lauson, H. D., Baldwin, E. de F., Pinchot, G. B., and Richards, D. W., Jr. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. VIII. Clinical use of concentrated human serum albumin in shock and comparison with whole blood and with rapid saline infusion. *Jour. Clin. Investigation* **23**: 491–505 (July).
10. Cronkite, E. P., Lozner, E. L., and Deaver, J. M. 1944. Use of thrombin and fibrinogen in skin grafting. *Jour. Amer. Med. Assoc.* **124**: 976–978 (April).

11. Dees, J. E. 1944. Chemical, clinical, and immunological studies on products of human plasma fractionation. XVIII. Fibrinogen coagulum as an aid in the operative removal of renal calculi. *Jour. Clin. Investigation* **23**: 576–579 (July).

12. DeGowin, E. L. 1944. Chemical, clinical, and immunological studies on the products of human plasma fractionation. XIV. Appraisal of isohemagglutinin activity. *Jour. Clin. Investigation* **23**: 554–556 (July).

13. a. Eagle, H. 1935. Studies on blood coagulation. *Jour. Gen. Physiol.* **18**: 531, 547, 813 (March).

b. Eagle, H., and Baumberger, J. P. 1935. Studies on blood coagulation: on constancy of hydrogen ion concentration during coagulation of fibrinogen by thrombin. *Jour. Gen. Physiol.* **18**: 809–812 (July).

14. Edsall, J. T., Ferry, R. M., and Armstrong, S. H., Jr. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. XVI. The proteins concerned in the blood coagulation mechanism. *Jour. Clin. Investigation* **23**: 566–572 (July).

15. Enders, J. F. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. X. The concentrations of certain antibodies in globulin fractions derived from human blood plasma. *Jour. Clin. Investigation* **23**: 510–530 (July).

16. a. Ferry, J. D., and Oncley, J. L. 1938. Studies of the dielectric properties of protein solutions. II. The water soluble proteins of normal horse serum. *Jour. Amer. Chem. Soc.* **60**: 1123–1132 (May).

b. Ferry, J. D., and Morrison, P. R. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. XVI. Fibrin clots, fibrin films, and fibrinogen plastics. *Jour. Clin. Investigation* **23**: 566–572 (July).

17. Fine, J., Frank, H. A., and Seligman, A. M. Traumatic shock. VIII. Studies in the therapy of hemodynamics of shock. *Jour. Clin. Investigation.* (In press.)

18. Flosdorf, E. W., and Mudd, S. 1935. Procedure and apparatus for preservation in "lyophile" form of serum and other biological substances. *Jour. Immunol.* **29**: 389–425 (Nov.).

19. Friend, D. G., Ferry, J. D., and Oncley, J. L. 1938. Dispersion of the dielectric constant of solutions of urinary proteins. *Jour. Biol. Chem.* **123**, Proc. XXXIX (March).

20. Green, A. A. 1938. The amphoteric properties of certain globulin fractions of normal horse serum. *Jour. Amer. Chem. Soc.* **60**: 1108–1115 (May).

21. Hawn, C. v. Z., Bering, E. A., Jr., Bailey, O. T., and Armstrong, S. H., Jr. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. XIX. A note on the use of fibrinogen and thrombin in the surface treatment of burns. *Jour. Clin. Investigation* **23**: 580–585 (July).

22. Heckel, G. P., Erickson, C. C., Yuile, C. L., and Knutti, R. E. 1938. Blood plasma proteins as influenced by the intravenous injection of gum acacia. *Jour. Exper. Med.* **67**: 345–360 (March).

23. a. Heidelberger, M. 1938. Protein constitution and immunological behavior. *Cold Spring Harbor Symposia Quant. Biol.* **6**: 369–375.

b. Heidelberger, M. 1939. Chemical aspects of the precipitin and agglutinin reactions. *Chem. Rev.* **24**: 323–343 (April).

24. a. Heyl, J. T., and Janeway, C. A. 1942. The use of human albumin in military medicine. Part I. The theoretical and experimental basis for its use. *U. S. Naval M. Bull.* **40**: 785–791 (Oct.).

b. Heyl, J. T., Gibson, J. G., II, and Janeway, C. A. 1943. Studies on the plasma proteins. V. The effect of concentrated solutions of human and bovine serum albumin on blood volume after acute blood loss in man. *Jour. Clin. Investigation* **22**: 763–773 (Nov.).

c. Heyl, J. T., and Janeway, C. A. Unpublished data.

25. a. Ingraham, F. D., and Bailey, O. T. 1944. The use of products prepared from human fibrinogen and human thrombin in neurosurgery. Fibrin foams as hemostatic agents; fibrin films in repair of dural defects and in prevention of meningocerebral adhesions. *Jour. Neurosurgery* **1**: 23–39 (Jan.).

b. Ingraham, F. D., Bailey, O. T., and Nulsen, F. E. Studies on fibrin foam as a hemostatic agent in neurosurgery with special reference to its comparison with muscle. *Jour. Neurosurgery* **1**: 171–181 (May).

c. Ingraham, F. D. and Bailey, O. T. Clinical use of products of human plasma fractionation. III. The use of products of fibrinogen and thrombin in surgery. *Jour. Amer. Med. Assoc.* (In press.)

26. a. Janeway, C. A. 1943. The plasma proteins: their importance in clinical medicine and surgery. *New England Jour. Med.* **229**: 751–756 and 779–785 (Nov. 11 and 18).

b. Janeway, C. A., Gibson, S. T., Woodruff, L. M., Heyl, J. T., Bailey, O. T., and Newhouser, L. R. 1944. Chemical, clinical, and immunological studies on the products of human plasma fractionation. VII. Concentrated human serum albumin. *Jour. Clin. Investigation* **23**: 465–490 (July).

c. Janeway, C. A. 1942. Immunological and clinical studies on purified proteins of human and animal plasma. Chapter XXI in *Blood Substitute and Blood Transfusions* by Mudd, S. and Thalhimer, W. Charles C. Thomas, Baltimore.

d. Janeway, C. A. Clinical use of products of human plasma fractionation: I. Albumin in shock and hypoproteinemia and II. γ-globulin in measles. *Jour. Amer. Med. Assoc.* (In press.)

27. a. Kabat, E. A. 1939. Molecular weight of antibodies. *Jour. Exper. Med.* **69**: 103–118 (Jan.).

b. Kabat, E. A., and Pedersen, K. O. 1938. Molecular weights of antibodies. *Science* **87**: 372 (April).

28. a. Longsworth, L. G., and MacInnes, D. A. 1939. Electrophoresis of proteins by the Tiselius method. *Chem. Rev.* **24**: 271–287 (April).

b. Longsworth, L. G., and MacInnes, D. A. 1940. Electrophoretic study of nephrotic sera and urine. *Jour. Exper. Med.* **71**: 77–82 (Jan.).

c. Longsworth, L. G., Shedlovsky, T., and MacInnes, D. A. 1939. Electrophoretic patterns of

normal and pathological human blood serum and plasma. *Jour. Exper. Med.* **70**: 399–413 (Oct.).

29. a. LUETSCHER, J. A., JR. 1939. Serum albumin. II. Identification of more than one albumin in horse and human serum by electrophoretic mobility in acid solution. *Jour. Amer. Chem. Soc.* **61**: 2888–2890 (Oct.).
    b. LUETSCHER, J. A., JR. 1940. Electrophoretic analysis of plasma and urinary proteins. *Jour. Clin. Investigation* **19**: 313–320 (March).
    c. LUETSCHER, J. A., JR. 1941. Electrophoretic analysis of proteins of plasma and serous effusions. *Jour. Clin. Investigation* **20**: 99–106 (Jan.).
    d. LUETSCHER, J. A., JR. 1944. The effect of a single injection of concentrated human serum albumin on circulating proteins and proteinuria in nephrosis. *Jour. Clin. Investigation* **23**: 365–371 (May).
30. LUNDGREN, H. P., PAPPENHEIMER, A. M., JR., AND WILLIAMS, J. W. 1939. The molecular weight of the diphtheria toxin protein. *Jour. Amer. Chem. Soc.* **61**: 533–534 (Feb.).
31. MACINNES, D. A., AND LONGSWORTH, L. G. 1939. Electrophoretic studies on blood sera. *Science* **89**: 438 (May).
32. MADDEN, S. C., AND WHIPPLE, G. H. 1940. Plasma proteins: their source, production, and utilization. *Physiol. Rev.* **20**: 194–217 (April).
33. a. MELLANBY, J. 1933. Thrombase—its preparation and properties. *Proc. Roy. Soc. London* sB **113**: 93–106 (June).
    b. MELLANBY, J. 1930. Prothrombase—its preparation and properties. *Proc. Roy. Soc. London* sB **107**: 271–285 (Dec.).
34. MORRISON, P. R., AND SINGER, M. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. XVII. A note on the absorption rates of fibrin films in tissue. *Jour. Clin. Investigation* **23**: 573–575 (July).
35. NEWHOUSER, L. R., AND LOZNER, E. L. 1942. The use of human albumin in military medicine. Part III. The standard Army-Navy package of serum albumin human (concentrated). *U. S. Naval M. Bull.* **40**: 796–799 (Oct.).
36. ONCLEY, J. L. 1940. Electric moments and relaxation times of protein molecules. *Jour. Phys. Chem.* **44**: 1103–1113 (Dec.).
37. ORDMAN, C. R., JENNINGS, C. G., AND JANEWAY, C. A. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. XII. The use of concentrated normal human serum gamma globulin (human immune serum globulin) in the prevention and attenuation of measles. *Jour. Clin. Investigation* **23**: 541–549 (July).
38. a. PAPPENHEIMER, A. M., JR., LUNDGREN, H. P., AND WILLIAMS, J. W. 1940. Studies on molecular weight of diphtheria toxin, antitoxin and their reaction products. *Jour. Exper. Med.* **71**: 247–262 (Feb.).
    b. PAPPENHEIMER, A. M., JR., LUNDGREN, H. P., AND WILLIAMS, J. W. Personal communication.
39. PUTNAM, T. J. 1943. The use of thrombin on soluble cellulose in neurosurgery; clinical application. *Ann. Surg.* **118**: 127–129 (July).
40. a. PILLEMER, L., ECKER, E. E., ONCLEY, J. L., AND COHN, E. J. 1941. The preparation and physicochemical characterization of the serum protein components of complement. *Jour. Exper. Med.* **74**: 297–308 (Oct.).
    b. PILLEMER, L. 1943. Recent advances in the chemistry of complement. *Chem. Rev.* **33**: 1–26 (Aug.).
    c. PILLEMER, L., ONCLEY, J. L., MELIN, M., ELLIOTT, J., AND HUTCHINSON, M. C. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. XIII. The separation and concentration of isohemagglutinins from group specific human plasma. *Jour. Clin. Investigation* **23**: 550–553 (July).
41. RUTSTEIN, D. D., GREENBERG, M., AND FRANT, S. Unpublished data.
42. a. SCATCHARD, G., GIBSON, S. T., WOODRUFF, L. M., BATCHELDER, A., AND BROWN, A. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. IV. A study of the thermal stability of human serum albumin. *Jour. Clin. Investigation* **23**: 445–453 (July).
    b. SCATCHARD, G., BATCHELDER, A. C., AND BROWN, A. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. VI. The osmotic pressure of plasma and of serum albumin. *Jour. Clin. Investigation* **23**: 458–464 (July).
43. SEEGERS, W. H. 1940. Purification of prothrombin and thrombin: chemical properties of purified preparations. *Jour. Biol. Chem.* **136**: 103–111 (Oct.).
44. STOKES, J., JR., MARIS, E. P., AND GELLIS, S. S. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. XI. The use of concentrated normal human serum gamma globulin (human immune serum globulin) in the prophylaxis and treatment of measles. *Jour. Clin. Investigation* **23**: 531–540 (July).
45. SVENSSON, H. 1939. Direkte photographische aufnahme von elektrophorese-diagrammen. *Kolloid-Z.* **87**: 181–186 (May).
46. WANGENSTEEN, O. H., KREMEN, A. J., AND STATE, D. Unpublished data.
47. WARREN, J. V., STEAD, E. A., JR., MERRILL, A. J., AND BRANNON, E. S. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. IX. The treatment of shock with concentrated human serum albumin: a preliminary report. *Jour. Clin. Investigation* **23**: 506–509 (July).
48. a. WILLIAMS, J. W., PETERMANN, M. L., COLOVOS, G. C., GOODLOE, M. B., ONCLEY, J. L., AND ARMSTRONG, S. H., JR. 1944. Chemical, clinical and immunological studies on the products of human plasma fractionation. II. Electrophoretic and ultracentrifugal studies of solutions of human serum albumin and immune serum globulins. *Jour. Clin. Investigation* **23**: 433–436 (July).
    b. WILLIAMS, J. W., AND PAPPENHEIMER, A. M., JR. Personal communication.
49. WOODRUFF, L. M., AND GIBSON, S. T. 1942. The use of human albumin in military medicine. Part II. The clinical evaluation of human albumin. *U. S. Naval M. Bull.* **40**: 791–796 (Oct.).

THE PRESENT STATUS OF PENICILLIN IN THE TREATMENT OF INFECTIONS

CHESTER S. KEEFER

Wade Professor of Medicine, Boston University School of Medicine

(*Read April 20, 1944, in the Symposium on Wartime Advances in Medicine*)

In 1929, Alexander Fleming wrote an essay for the *British Journal of Experimental Pathology* entitled "An Antibacterial Action of Cultures of Penicillium with Special Reference to Their Use in Isolation of B. Influenzae." This paper marked the beginning of one of the most remarkable discoveries in modern medicine. Since its contents attracted so little attention for such a long period of time, I propose to introduce my subject today by reviewing some of the historical landmarks in the development of penicillin as an antibacterial agent.

At the outset it should perhaps be stressed that Alexander Fleming was not searching for a practical therapeutic agent when he first discovered penicillin. In fact, the mold which contaminated his culture plates threatened to ruin his original studies. The real purpose of his investigation was to study the degradation of colonies of staphylococci as they aged under conditions of artificial cultivation. As his work progressed, he noticed one day that a colony of mold was growing on one of his culture plates. Instead of discarding this experiment he continued to watch the growth of the mold and the colonies of staphylococci, side by side. With the passage of time a very curious phenomenon excited his interest and curiosity. As the mold increased in size, the colonies of staphylococci in the neighborhood began to decrease in size and finally disappeared. This observation led Fleming to investigate the mechanism of this reaction. He isolated the mold, and after growing it for seven to ten days on the surface of a liquid medium, he found that the broth contained a substance that had a powerful effect in inhibiting the growth of many species of bacteria. He named the substance penicillin. It was remarkable that this material had a high degree of specificity in spite of its broad action.

Many of the gram positive cocci were unable to survive and reproduce in the presence of penicillin but the gram negative bacilli were not only able to survive but also to flourish. Fleming did not stop here. He discovered that this agent was non-toxic for animals, and that the colorless corpuscles could withstand exposure to comparatively high concentrations without being killed. It was non-irritating to the human skin and mucous membranes. There were, however, certain obstacles in the way of producing this substance in large quantities and in a stable form so that it might be used for the treatment of infections in man. These various characteristics were fully recorded and the suggestion was made that penicillin might be developed as a therapeutic agent. For many reasons, this suggestion did not become an accomplished fact for some years. However, in spite of the failure to produce large quantities of penicillin, Fleming recommended its use as an aid to the bacteriologist in isolating gram negative organisms from mixed cultures. Thus, the bacteriologist had a new method for suppressing the growth of gram positive organisms in mixed cultures, so that gram negative bacilli could grow freely.

Very little interest was shown in this subject for a period of ten years, when Professor Florey and his associates at Oxford University undertook the problem of producing enough penicillin for the treatment of a small number of patients with staphylococcic infections. They were successful, and the results which were obtained in the first few cases were sufficiently promising to encourage them and others to develop this field of investigation further. Florey was searching for an agent that could be used in the treatment of wound infections. He knew, that one of the common micro-organisms in wounds was the hemolytic staphylococcus aureus, and he was also familiar with the fact that the sulfonamides had little or no effect in this group of infections. From the early experience with penicillin there were good reasons for believing that it had all of the characteristics of a great antibacterial agent. It was non-toxic, it was powerful, and it was effective in influencing the course of human infections in a favorable direction.

There were many stumbling blocks, however, in the way of obtaining large amounts of penicil-

lin. The yield of active material from the cultures was small; it was extremely unstable in its crude form; there were great losses in the recovery process; and the finished product had to be carefully treated in order to prevent loss of potency.

After a visit to the United States by Professor Florey, a number of investigators began to study penicillin. Many of these studies were initiated and stimulated by one of your distinguished members, Dr. A. N. Richards, Chairman of the Committee on Medical Research of the Office of Scientific Research and Development. In a recent statement, Dr. Richards has outlined in a clear-cut and concise manner the history of the development of penicillin in the United States. He has pointed out that the entire program was started and has been continued as a part of the war effort. It is fair to say it is one of the outstanding wartime advances in medicine.

The report today is based on my personal observations of a considerable number of patients who have received penicillin, and on a study of the records of several thousand cases collected by the Committee on Chemotherapeutic and Other Agents of the National Research Council. It is no exaggeration to say that penicillin is nontoxic. It will save many lives and reduce many days of acute and chronic illness. The material itself is extremely soluble in water, in physiological saline solution, and in solutions of dextrose.

It is inactivated by gastric juice and poorly absorbed from the intestinal tract so that it cannot be given by mouth. It can be administered in solution, intravenously, intramuscularly, or it can be applied locally to wounds. Once it enters the blood stream, it is rapidly excreted by the kidney into the urine and by the liver into the bile. We have not found it in the tears, saliva, gastric juice, succus entericus or cerebrospinal fluid. It is absorbed slowly from the pleural cavity and the subarachnoid space, and it is absorbed more slowly from the subcutaneous tissue than from the muscles. In order to obtain the best results, therefore, it is necessary to give penicillin by repeated intramuscular or intravenous injections or by continuous intravenous infusions.

STAPHYLOCOCCIC INFECTIONS

Penicillin is the best therapeutic agent available for the treatment of staphylococcic infections with and without bacteremia. The fatality rate can be reduced, and the total number of days of illness greatly shortened. Of necessity, it must be combined with good surgical management in the treatment of surgical infections. It can be used as a preoperative as well as a postoperative measure to aid in the prevention of bacterial invasion and in the promotion of conditions which favor wound healing. Penicillin also aids in the localization of infections. In many cases it helps greatly in the healing process. When these infections occur in patients with diabetes or in persons with an inadequate circulation to the infected part, then larger quantities are necessary and the problem is more complex. When the infection is localized on the heart valves, then the outlook is extremely grave in spite of penicillin.

PNEUMOCOCCIC INFECTIONS

The pneumococcus is extremely sensitive to the action of penicillin. It is effective in pneumococcus pneumonia, in infections of the pleura, usually when assisted by surgical drainage. It has shown promise in cases of pneumococcic meningitis and even in a few cases of endocarditis.

We have much to learn in the treatment of pneumococcus meningitis, but when one recalls that before the use of the sulfonamides, very few patients with such an infection ever survived, it is extraordinary that with modern methods of treatment more than half of those infected are now being saved.

HEMOLYTIC STREPTOCOCCIC INFECTIONS

Like the pneumococcus, the hemolytic streptococcus is extremely sensitive to the action of penicillin. Indeed one of the first cases treated with penicillin in this country by Dr. Francis Blake was a case of hemolytic streptococcus infection which had been resistant to the sulfonamides. In some cases of infection the action is so striking and dramatic that recovery is prompt. The organisms either disappear or decrease in number from the primary focus. If treatment is stopped too soon, there is often a return of the organisms in the focus and a recurrence of the infection. This is a question which needs further study and investigation. However, it is plain that hemolytic streptococcus infections can be controlled with penicillin.

GONOCOCCIC AND MENINGOCOCCIC INFECTIONS

These two organisms are the only gram negative cocci which are sensitive to penicillin. In

fact, they are more sensitive to penicillin than any other organisms. Localized gonococcic infections of the urethra can be cured within twenty-four hours with as little as 100,000 units of penicillin. Infections localized elsewhere, as in the heart valves, are much more resistant to treatment and are sometimes not influenced favorably. These preliminary observations once again emphasize the differences that exist in the treatment of infections depending upon their location.

The treatment of meningococcus meningitis is still being explored. The sulfonamides have been so outstandingly successful that very few cases have been treated with penicillin.

OTHER DISEASES

There are good grounds for saying that penicillin will be an extremely useful and good agent in the treatment of syphilis. The treponema pallidum is very sensitive to its action and experiments are under way at present to define dosage and time relationships in the treatment of this disease. The results are extremely promising and we should know much more about the efficiency of this chemical in the treatment of syphilis in due course of time.

I have already said that the location of an infection often influences its seriousness in spite of the fact that the infecting organism may be susceptible to the chemotherapeutic agent. This is brought out in a striking way in the study of bacterial endocarditis. Here is a disease in which the infecting organism is usually susceptible to the action of penicillin, and yet the administration of it is often followed by most disappointing results. One can frequently clear the blood of organisms, and the temperature decreases. Temporary remissions may occur and periods of improvement may follow its use. It is too early to make any statements concerning the exact position of penicillin in this disease.

DISEASES IN WHICH PENICILLIN IS NOT EFFECTIVE

No discussion of a new therapeutic agent would be complete without saying something about the diseases in which it is ineffective.

We know from experience that this antibacterial agent is of no value in the gram negative bacillary infections and it does not influence virus infections. It is ineffective in tuberculosis, acute rheumatic fever, Hodgkins disease and leukemia, malaria, histoplasmosis, toxoplasmosis, coccidiomycosis and blastomycosis, ulcerative colitis, pemphigus vulgaris, and disseminated lupus erythematosis. Penicillin has followed the rule that, if it is ineffective in the test tube or in the experimental animal, it is ineffective in man.

I have said that penicillin is non-toxic. There have been a few side reactions following its use which require comment. None of these reactions have been serious and all of them have been either of a temporary or transient character. They have consisted of nausea, headache, tingling of the hands and feet, thrombophlebitis, urticaria, fever, or chills and fever. In no case has any serious reaction occurred, and no permanent damage to tissues has ever been recorded.

In summing up, then, it can be said that penicillin is a remarkable antibacterial agent. It has a powerful effect on many bacteria, and it is nontoxic. It will be responsible for saving many lives and reducing the total number of days of acute and chronic illness. Since the normal habitat of *Penicillium notatum* is in the soil, one can say that another good thing has come out of the earth.

RECENT ADVANCES IN THE TREATMENT OF WOUNDS

ALLEN O. WHIPPLE

Valentine Mott Professor of Surgery, Columbia University

(*Read April 20, 1944, in the Symposium on Wartime Advances in Medicine*)

THE repair of a loss of continuity of tissues—a wound—is a complex biological process which begins with the injury and ceases when the continuity of the tissues is restored. It is a summation of growth inhibiting and growth promoting factors. There are two phases in this wound healing phenomenon, the latent or lag period during which time tensile strength is in abeyance, and which is prolonged by certain factors, and a period of fibroplasia when, with those factors removed, tensile strength steadily increases to the completion of the union. This natural phenomenon of wound healing increases in its complexity in the ascending scale of the species becoming most involved in man with his efforts to assist nature in the repair of the wound and accelerate the rate of healing.

From time immemorial man has sought a magic substance that placed in or on a wound, regardless of inhibiting factors, will hasten the healing. And today the search for this wound healing factor goes on apace, unmindful of the factors that interfere with the normal natural process in its two phases. Uncontrolled experiments and inaccurate clinical observations and impressions continue to mislead so-called investigators into believing they have discovered this magic stimulant. It has not been discovered. The search for such a substance should continue as an adjuvant to the sound surgical principles of wound repair which have now been established, but this search should be made with a full appreciation of the inhibiting factors that interfere with the natural processes of repair.

The inhibiting factors that delay or prevent wound healing and that must be eliminated in treating wounds are:

First Factor: *Interference with the Blood Supply and the Nutrition of the Injured Tissues.*—This may be the result of damaged or severed vessels, to ischemia of wound surfaces and adjacent tissue as a result of bad surgery, or it may be due to an anemia from previous or continued blood loss or sepsis. In the case of damaged vessels, paravertebral block may improve the circulation by relaxing the vasospasm. If the main arterial or venous trunk is severed or torn, new methods of non-suture anastomosis and vein graft have been devised to restore the circulation to the injured limb. With blood loss and anemia, by restoring blood and hemoglobin volumes, lives are saved that were formerly considered hopeless.

Second Factor: *Infection.*—Every accidental wound is contaminated and potentially infected by bacteria introduced by the agent producing the injury or by the air and particulate matter entering the wound after injury. During the first six to twelve hours, the lag period or period of contamination, bacteria on the surfaces of the wound are inactive. If the wound is properly treated during this time, by removing all nonviable and dead tissue, blood clot and foreign bodies, irrigating the wound from the bottom of the wound outward with normal salt solution and leaving clean fresh surfaces, the wound will heal normally without infection. If the wound edges are approximated without tension and immobilized, the healing period will be shortened by primary union. Such clean wounds, free of foreign body with wound surfaces approximated, have an amazing capacity for combating contamination, whereas blood clot, dead and non-viable tissue and foreign bodies, which cannot be reached by blood and immune body fluids and phagocytes, if left in the wound, particularly favor bacterial growth and invasive infection.

The use of the bacteriostatic agents, the sulfonamides and penicillin, is the new phase of wound therapy which has accomplished results never before realized. These agents, when combined with sound surgery, have reduced invasive infection and sepsis to an amazing degree and in the case of penicillin offer unheard of possibilities in combating certain bacteria.

With the lay public and in the minds of many of the medical profession these bacteriostatic agents are looked upon as miracle drugs and a great misconception exists as to their efficacy and their mode of action. Too little attention is paid to the differences between so-called medical

and surgical infections, and the use of these bacteriostatic agents in the therapy of these two types of infection. In most of the infections treated medically there is no marked disturbance in the circulation in the inflamed organ or tissues and there is no breakdown of tissue in the form of abscess or sequestration. In such infections the chemotherapeutic agent reaches the inflamed tissues and cells in the interstitial fluid through the intact circulation. The bacteria are inhibited in their growth and the damaged bacteria, distorted and arrested in division, become easy prey to the phagocytes of the circulating blood and reticulo-endothelial cells and to the lytic fluids and immune bodies in the intercellular fluid.

In surgical infections,[1] that is infections requiring operative therapy, the process is quite different. In such infections cell and tissue destruction takes place as a result of trauma or by direct bacterial action or because of thrombosed vessels or the pressure necrosis of a developing abscess. Lysis of cells with pus formation and tissue sequestration develop. Into such areas with obstructed circulation the bacteriostatic agents cannot penetrate to reach the bacteria propagating in the necrotic tissue, pus and sequestrum and the all important process of phagocytosis ceases to function beyond the zone of tissue bordering on the abscess. Not unless the abscess is small or near the surface or until the abscess is incised and drained, pus is evacuated and sequestrated dead tissue is removed, with the restoration of circulation to the inflammatory zone, can the bacteriostatic agents reach the bacteria and phagocytes and immune bodies function to eliminate them. Therefore, in surgical infections these agents will always remain as adjuvants or adjuncts to sound operative procedure. But they are of the greatest help and in some cases the only means of preventing the systemic spread of infection beyond the local lesion and in combating it as a septicemia and bacteriemia.

Third Factor: *Foreign Bodies.*—By this is meant any inert object introduced into the wound or developing in the wound as a result of tissue necrosis. Bacteria in a sense are foreign bodies but they are as a rule active with a tendency to invade the tissues on which they are propagating. Foreign bodies usually introduce bacteria into the wound, some much more threatening than others, such as soiled clothing and fertilized earth. The bacteria, aerobic and anaerobic, on such foreign bodies multiply rapidly after the latent hours of contamination and are not in contact with circulating blood or body fluids so that chemotherapeutic agents cannot reach them and they continue to be a nidus of infection in the wound until removed surgically. Necrotic and sequestrated tissue, especially loose bone that is not dissolved by lysis, behaves in the same way and prevents wound healing, regardless of chemotherapy, until removed from the wound.

Fourth Factor: *Lack of Immobilization.*—Motion in the wound edges and surfaces prolongs the lag period by increasing bleeding and exudate and by interfering with fibroplasia and fusion of wound surfaces. It is in this regard that proper surgery can assist nature most effectively by approximating wound surfaces and immobilizing the parts by adequate hemostasis, suturing and splinting.

Fifth Factor: *Fluid, Salt, Protein and Hemoglobin Imbalance.*—Blood chemistry and metabolism studies have contributed very significantly to our understanding of the causes of delayed wound healing especially in the established infections in compound fractures and in extensive third degree burns. The disproportionate increase in interstitial fluid with decrease in blood and hemoglobin volumes, along with the depletion of protein stores that have been recently demonstrated[2] in these severe established infections, require four essential therapeutic measures.

(1) Control of the infection by bacteriostatic agents. Of these penicillin has proven the most effective because of its far greater potency, its control of pyogenic bacteria resistant to the sulfonamides, its activity in the presence of pus and wound exudate, and its non-toxicity in maximum therapeutic doses. Until the bacterial infection in the zone surrounding the infection is controlled, the other three measures will be ineffective.

(2) Replenishing the depleted blood and hemoglobin volumes by whole blood transfusions. This is essential before operative procedures are undertaken because of the relative anoxia of the tissues. Until the anoxia is corrected, these patients may go into severe shock with even minor surgical procedures, such as skin grafting.

[1] Meleney, F. L. The Difficulty of Evaluating Drug Treatment in Surgical Infections. *Jour. Amer. Med. Assoc.* **124**: 1021, April 8, 1944.

[2] Lyons, C. Penicillin Therapy of Surgical Infections in the U. S. Army. *Jour. Amer. Med. Assoc.* **123**: 1007, December 18, 1943.

(3) The third factor is revision of the wound to remove sequestra and to drain pockets of pus. Compound fractures will not heal as long as bone sequestra remain in the wound. Similarly third degree burns require the removal of skin sloughs and early skin grafting, but usually the anemia, hypoproteinemia, and fluid and salt imbalance must be corrected before skin grafting can be successfully done. Following surgery, penicillin given locally and systemically, these wounds and burns heal amazingly.

(4) High protein diet and split protein hydrolisates systemically are needed to correct the nitrogen imbalance, restore lost protein, and to build up the muscle atrophy and weight loss in these patients.

It is the factor of contamination and infection and the role of the recently developed bacteriostatic agent penicillin as an adjuvant to sound surgery that I wish to discuss at this time.

The period of inactivity of contaminating bacteria can be prolonged by the introduction into the wound or into the circulating blood and outflowing exudate, of a bacteriostatic substance. The onset of systemic infection during the period of contamination can be prevented by enlarging the wound to expose all surfaces, to remove all devitalized and dead tissue and foreign bodies and to provide free drainage of wound exudate, and by placing the wound at rest by immobilization. If infection has begun, that is if bacteria have begun to grow on the wound surfaces and in the retained contents and to invade the deeper structures, it can be arrested by opening the wound to remove foreign bodies and dead tissue and provide free drainage, and the spread of infection can be prevented by a bacteriostatic agent given systemically so that the interstitial fluids containing this bacteriostatic agent will come in contact with the invading bacteria, stop their growth, and permit the phagocytes to dispose of the impotent organisms.

To understand the role of these bacteriostatic agents in surgical infections one must appreciate their mode of action and their activity in relation to lesions requiring surgical therapy.

The difference between a bacteriostatic agent and a bacteriocidal agent is that the former arrests the metabolic processes and division of the bacteria, the latter, as a protoplasmic poison, kills the organisms. Gardner[3] was the first to describe the morphological changes which take place when sensitive bacteria are grown in a dilution of penicillin, insufficient to cause complete inhibition of growth. By this interference with division and multiplication, bizarre giant forms are seen and in the case of bacilli and vibrios long filaments appear. But the respiration of bacteria, as measured by the Warburg-Barcroft apparatus, is not affected by strong solutions of penicillin as compared to most antiseptics which, acting on some protoplasmic constituent, rapidly stop respiration.

The question naturally arises as to why the bacteriostatic agents inhibit gram-positive organisms so much more than the gram-negative varieties. Dubos[4] has recently discussed this subject and I quote this well-known authority. "Very few agents are capable of acting on the gram-negative bacilli, whereas the list of those active against the gram-positive species comprises a large variety of substances of many different chemical types. In general it may be said that practically all agents which are active against gram-negative bacilli behave as protoplasmic poisons, whereas several of the substances which are active against the gram-positive species are little toxic for tissue cells. Thus it would appear that the gram-negative organisms are much closer to the mammalian cells than are the gram-positive.

"There is evidence that the protoplasm of the gram-positive organisms is characterized by an overall isoelectric point much more acidic than that of the gram-negative bacilli and of the tissue cells. There is also evidence that the cell walls of the two groups of organisms differ in structure and behavior and it is likely, therefore, that permeability properties may account, in part at least, for the differential susceptibility to bacteriostatic agents."

These observations would seem to establish the rationale that to interfere with the pathogenic function of bacteria it is not necessary to affect their viability so much as to alter their potential damage to the host, in this case their ability to multiply, and to render them vulnerable to the phagocytes and other protective body mechanisms. A bacteriostatic agent may be effective on a surface during the lag period of bacterial growth by prolonging the period of contamination, but if infection has become established and invasion of tissues by the multiplying bacteria

[3] Gardner, A. D. *Nature* **146**: 837, 1940.

[4] Dubos, R. In Symposium on War Surgery, Meeting of the Association of War Surgeons, October, 1943, in *Clinics*. Lippincott, 1944.

has begun, the agent must be delivered by the circulating blood to the interstitial and intercellular fluid to come in contact with the invading bacteria, to arrest their growth, and to make them vulnerable to the leucocytes and phagocytes of the invaded area.

It is obvious that, if bacteria are segregated in foreign bodies or in a bone or dead tissue sequestrum, the body fluids cannot deliver the bacteriostatic agent. The same is true in an undrained walled off abscess, isolated by thrombosed vessels and abscess membrane from circulating blood. And it is just this isolation and sequestration that differentiates a medical lesion from one requiring surgical therapy. In a medical lesion, such as a pneumonia, there is no breakdown of tissue with abscess formation and sequestration of tissue. The circulation is not obstructed and a bacteriostatic agent can reach all parts of it. In a lesion that has tissue breakdown, abscess formation and sequestration, and extension of the devitalized area by bacteriolysis and increased tension of the inflammatory zone, incision and drainage and removal of sequestrated tissue or foreign body is essential. For until this is done the restoration of the circulation and subsidence of the inflammatory process will not take place, and a bacteriostatic agent can do no more than prevent a systemic invasion. But with adequate and correct surgery the bacteriostatic agent can be introduced into the wound as well as into the circulating blood to reach the subsiding inflammatory zone. All of which adds up to what should be obvious to every thoughtful surgeon and that is that the bacteriostatic agents are a wonderful adjuvant, but only an adjuvant, to correct and adequate surgery.

PENICILLIN

Penicillin is an unstable acid produced by extraction from mass cultures of the fungus *Penicillium notatum*. The preparations used therapeutically are its salts, the sodium salt usually employed for systemic treatment, the calcium salt for local applications. It is fortunate in many ways that when first tried therapeutically by the group of scientists in Oxford it was obtainable only in small amounts and was demonstrated the hard way by the experimental method. Before any trial was made clinically, it had been studied[5] by its discoverer Andrew Fleming (1929) in his bacteriological laboratory, and some ten years later it was tried out by Dr. Florey[6] and his associates in the whole range of infections. It was demonstrated to be non-toxic in maximum therapeutic doses, far more potent than any other known bacteriostatic agent against many organisms that were resistant to the sulfonamides, and was not inhibited by pus and para-aminobenzoic acid. It is not injurious to body tissues or cells in competent therapeutic doses, and from the standpoint of accidental and war wounds, it is the most effective of the bacteriostatic agents against the three most serious infective organisms, the streptococcus, the staphylococcus, and the anaerobes.

It is now being produced in increasing amounts at decreasing cost in North America and Great Britain. According to a recent statement by Surgeon General Kirk[7] in December the Army received 56 percent of the total supply in the United States, the Navy 18 percent, the U. S. Public Health Service 2 percent, the Office of Scientific Research (for civilian use) 15 percent, and the manufacturing firms retained the remainder, 9 percent, for their own research. Though production is mounting steadily the supply is below the demand in all the government agencies. The proportion allocated to civilians is still limited to clinical research. Compared to the prices effective in February, 1943, the present cost of penicillin per 100,000 Oxford units is now approximately one-fourth that of a year ago.

The use of penicillin is not a simple and fool proof procedure. There are many points about the therapy that are still poorly understood and experimental. Although the preparations we are now getting are much purer and more potent, there are still many impurities in the product, and until it is produced synthetically in pure form, the possibilities in its use will be undetermined. It has been produced in crystalline form which should help in establishing the unitage and metabolism of the agent.

There are many bacteria, especially the gram-negative group, that are resistant to penicillin and some of the gram-positive organisms considered most susceptible show individual degrees of resistance. Before continuing the administration of penicillin the resistance of the organism recovered from the blood or the lesion should be

[5] Fleming, A. *Brit. Jour. Exp. Path.* **10**: 226, 1929.

[6] Abraham, E. P., Chain, E., Fletcher, C. M., Florey, H. W., Gardner, A. D., Heatley, N. G., and Jennings, M. A. *Lancet* **2**: 177, 1941.

[7] Kirk, N. T. *Jour. Amer. Med. Assoc.*, **123**: 974, 1943.

determined, and this is easily done in twenty-four hours by the trench method.

It is now well established that the susceptible organisms show group variations of susceptibility. In the order of susceptibility and the dosage of penicillin required to inhibit their growth the organisms may be listed as follows: gonococcus, meningococcus, pneumococcus, hemolytic streptococcus (A), staphylococcus aureus, streptococcus viridans, staphylococcus albus, and gram-positive rods. Mixtures of organisms, that is a susceptible staphylococcus with resistant proteus or pyocyaneus, will require a larger dosage to eliminate the susceptible coccus.

Another important factor in determining the appropriate dosage is the type of lesion present. The diffuse and phlegmonous lesions respond much more promptly than the localized lesions. The reasons for this have been discussed.

It is also recognized that in the initial stages of treatment maximal dosage should be used to give a knockout blow to the invading or circulating bacteria, for it is now known that certain strains of staphylococci build up a resistance to penicillin in small doses.

It is well established that the systemic administration of penicillin is the most effective method. But there is still considerable debate as to the best way of giving it systemically. The intermittent intravenous route has the objection that it has to be given every three to four hours and most of it is rapidly excreted by the kidneys. By continuous drip infusion the rate of excretion is slowed down, but because of the impurities, thrombophlebitis of the veins is common. The subcutaneous and the intramuscular injections are so painful for many patients that it becomes a serious problem. The majority of the investigators have been using it intramuscularly. It has been demonstrated that penicillin does not pass from the blood to the meninges and joints and pleura as readily as into other tissues. For this reason the agent should be injected as a solution, intrathecally and into the serous cavities.

At the Presbyterian Hospital we have used the intramuscular route in the majority of some 200 cases. The sodium salt is readily soluble in isotonic saline and can be conveniently administered in amounts of 5,000 units per cc. If large doses are required, the concentration may be increased to 10,000 units per cc.

{5,000 to 25,000 units every 3 to 4 hours,
{180,000 to 2,250,000 per course.

For local application solutions containing 250 to 1,000 units per cc. have been used, but this depends on the organism recovered. Doctors Meleney and Pulaski have developed a carbowax ointment that gives an even penicillin diffusion and effective in this ointment form up to sixty days.

We have not as yet definite data regarding the use of penicillin as a prophylactic measure in compound wounds, especially compound fractures, nor have we sufficient data on the use of penicillin in the prophylaxis of burns and the use of it in established infections in burns. Six centers are studying these problems at present.

The most important study of the local use of penicillin in established surgical infections is that reported by Dr. M. E. Florey[8] in a series of 100 hand infections with 100 controls treated by her in the Accident Infirmary in Birmingham, England.

Biweekly bacteriological examinations showed that in 75 percent of control wounds pyogenic organisms persisted from the first incision until final healing. The remaining 25 percent lost their infection or showed a diminished growth only when healing was far advanced. In 95 percent of the penicillin treated cases these organisms had markedly diminished in numbers or had disappeared before the end of the first week.

	Control	Penicillin treated
1.	Continuance of pain and throbbing for variable periods after incision.	Almost invariable relief of pain after incision and application.
2.	Pus continued until healing well advanced.	Pus scanty or absent.
3.	Excessive granulation with subsequent contraction and limitation of movement and sometimes painful tender scars. In severe infections usually much sloughing, death of tendon and bone and loss of digits.	Little granulation. Rapid and more complete restoration of function and smooth scars. No loss of phalanges or any other tissue when penicillin treatment was begun at first operation.

The working time saved to the 100 penicillin treated cases is said to total in the aggregate a number of years.

This review of the subject will, I hope, leave no uncertainty as to the role of penicillin in surgical lesions and infections. It is the most potent adjuvant to sound surgery of any of the bacteriostatic agents but it is only an adjuvant and can never be more than an adjuvant to sound surgery.

[8] Florey, M. E., and Williams, R. *Lancet* **1**: **73**, 1944.

THE ROLE OF IMMUNIZATION IN WARTIME ADVANCES IN MEDICINE

R. E. DYER

Assistant Surgeon General, U. S. Public Health Service; Director, National Institute of Health

(*Read April 20, 1944, in the Symposium on Wartime Advances in Medicine*)

INSOFAR as the effectiveness of troops is concerned, it matters little whether a casualty is occasioned by machine guns, bombs, artillery, or micro-organisms. Except in case of death, it is the time lost by the individual from effective action and the time of other individuals required for the care of the casualty which are important. The casualties among troops caused by some diseases are much greater than by other diseases just as the effectiveness of the modern machine gun is greater when measured by resulting casualties than is that of the rifle. The diseases of the greatest military importance are those of a communicable nature which occur in epidemics, whether the method of communication is by person to person contact, through the agency of food and water, or through the intervention of an insect. In the past wars the diseases which have caused the greatest damage to armed forces have been typhoid, dysentery, typhus, the venereal diseases, influenza, pneumonia, and, in the Spanish-American war, yellow fever. Measles, mumps and meningitis also have been important. Standing a little apart from these are two diseases which do not initiate the casualty but increase the death rate or time off duty among the casualties caused by enemy action—tetanus and gas gangrene.

The diseases mentioned above are diseases of the temperate zone and the reason our concern has been chiefly with the impact of these diseases on troops is because the great wars in which we ourselves have been interested, or in which our forebears of the past twenty centuries have been interested, have been fought in the temperate zones. Our share of the present war has so far been fought in the tropics and subtropics which fact has brought the so-called tropical diseases into a position of greater importance, particularly malaria as an actuality, and yellow fever as a threat. Other diseases to which those in charge of the health of our troops must give thought are plague, cholera, and tsutsugamushi or scrub typhus, dengue, and filariasis. There are additional diseases of a communicable nature, occasional cases of which may occur among our troops, but the diseases mentioned above will encompass most of our disease casualties. Smallpox has not been listed as a disease of military importance since the use of smallpox vaccine has long since eliminated it from consideration as a casualty-producing disease.

Prior to the last war we had for the active immunization of our troops vaccines against typhoid, cholera, and plague. Antitoxin which confers passive immunity of short duration against tetanus was available also. We had no agents of any proven preventive value against the venereal diseases, influenza, typhus, malaria, yellow fever, dengue, filariasis, pneumonia, meningitis, measles, and mumps. In the present war we are better prepared although no definite advance has been made in the field of active immunization against many of these diseases. The sulfa drugs have reduced the terrors of pneumonia and meningitis, the use of convalescent serum has proven of value in a somewhat restricted way in passive protection against measles, and the sulfa drugs and penicillin are revolutionizing the treatment of two of the venereal diseases. Plague and cholera vaccines are probably somewhat improved over the preparations used twenty years ago, while a vaccine against influenza is under scientific test at present.

There are, however, four important wartime diseases against which we have made recent and definite progress in the methods of active immunization of our troops, namely, typhoid fever, tetanus, yellow fever, and typhus fever.

IMPROVEMENT OF TYPHOID VACCINE

It was shown in Germany in 1896 that animals could be immunized against experimental infection with typhoid by inoculation with killed typhoid bacilli followed by graded doses of virulent organisms. Volunteers were then inoculated with a suspension of killed organisms and although a rather severe reaction of short duration followed, it was found that such a procedure was

relatively safe. At approximately the same time, and independently, similar experiments were carried out in England. Such a vaccine was used to protect British troops in the Boer War. Briefly, the method of manufacture of typhoid vaccine then employed entailed the growth of the micro-organisms on media, the killing of the harvested bacilli by heat and the addition of a small amount of phenol as a preservative. There have been some modifications in the techniques used at that time but the same general methods of manufacture have been used up to the present. The first attempts to use this method of immunization in this country were begun by the United States Army in 1908 utilizing the strain of typhoid bacillus employed by the English and known as the Rawlings strain. This strain was recovered by the British from a fatal case of typhoid fever in 1900. The choice of this strain was based on the fact that it produced less severe reactions than other strains considered and organisms harvested from solid media were easily emulsified. Little was known about methods of potency testing at that time but the inoculation of this strain stimulated the production of at least one demonstrable type of antibody, agglutinins, in an adequate manner.

Subsequent to the investigation carried out by the United States Army, typhoid vaccine was gradually introduced for the protection of civilians. This same Rawlings strain of typhoid bacillus was used for the manufacture of vaccine by the commercial producers who employed essentially the same methods as used by the United States Army.

During the first world war paratyphoid A and B strains were added to the typhoid vaccine making the so-called triple vaccine. Since the incidence of paratyphoid infections among troops in this country is low and the addition of paratyphoid strains to the vaccine increases the severity of reactions the paratyphoid strains were dropped at the close of the last war. With the advent of the present war the paratyphoid strains were again added to the vaccine given to troops because of their expected service in foreign lands.

The efficacy of typhoid vaccination has been widely discussed since its introduction as a compulsory measure by our Army in 1910. For the purpose of this paper it will be unnecessary to discuss the part that vaccination played in reducing the typhoid rate in the Army as against the part played by better sanitation. Improved sanitary measures undoubtedly played their part but could hardly be relied upon to explain the sudden drop in the typhoid rate for our troops from 335 per 100,000 in 1909 and 243 per 100,000 in 1910, to 85 and 30 in 1911 and 1912. In the Spanish-American war the rates per 100,000 of our troops for the years 1898 and 1899 were 14,158 and 2,064, while the rates for 1918 and 1919 were 30 and 46 respectively and those at present are much lower.

In 1934 the staff of the Army Medical School undertook a series of investigations in search for a strain of typhoid bacillus which would be antigenically superior to the Rawlings strain. This investigation was stimulated by the discovery of several investigators that different strains of typhoid organisms differed in antigenicity and that these variations were determined to some extent by the virulence of the organism. It had been demonstrated that mice immunized by vaccines made from virulent strains of typhoid organisms show a greater degree of immunity to subsequent challenge doses than do mice which have been immunized with vaccines prepared from less virulent strains.

The virulence in mice of seven strains was studied by the Army investigators—four of these strains were variants of the original Rawlings strain and three were freshly isolated strains; one isolated from a clinically mild case of typhoid, one from a clinically severe case, and one from a chronic carrier in Panama. Three of the variants from the Rawlings strain were of low virulence, while the fourth Rawlings strain had high virulence due to preliminary repeated passages through mice. The strain isolated from the mild case was of low virulence, while the strains from the severe case and the chronic carrier were of high virulence. Vaccines were then prepared from these seven strains and a separate group of mice immunized with each vaccine. Each group was then subdivided into seven groups and each challenged with a different strain. It was found that the three virulent strains produced a definitely higher degree of protection than the strains of low virulence. It was further found that the sera of volunteers immunized with vaccines made from strains of high virulence afforded a higher degree of protection to mice when tested for passive immunity than did vaccines made from less virulent strains. As a result of this work the Panama carrier strain known as Strain No. 58 has been substituted for the preparation of all Army typhoid vaccine and is now generally used in commercial production.

The scientific evidence justifies us in thinking that this study of typhoid strains has resulted in a definite improvement in typhoid vaccine. Just what influence this will have in the further lowering of morbidity and mortality rates remains to be determined.

THE DISCOVERY OF TETANUS TOXOID

The tetanus bacillus when cultivated under suitable conditions produces a toxin which is lethal to animals, including man. Spores are formed by this bacillus and these spores remain viable for long periods of time. When tetanus spores are placed in suitable media, or when their surroundings become favorable, bacilli develop, multiply, and produce the lethal toxin. The presence of other organisms in addition to tetanus spores in wounds, or the action of other agents such as the presence of dead tissue, bring about conditions which are favorable for multiplication and toxin production. Lacerated and puncture wounds received in battle are particularly suited for the development of this bacillus. Deposited in clean wounds, tetanus spores may lie dormant for some time only to develop when some secondary agent changes the local tissue reaction. For this reason a satisfactory immunizing agent for the prevention of tetanus must remain active as long as the spores are present. Such an agent has been developed since the last war.

It was found in the early nineties that a single injection of a sublethal amount of tetanus toxin into an animal results in the production of antibodies by the tissue cells of the body and that these immune bodies are present in the blood serum. An animal so treated is then actively immune to subsequent injections of toxin, and the blood serum of this immune animal if given to a second non-immunized animal confers a passive or borrowed immunity on the second animal. This passive immunity persists until the foreign antitoxic serum is eliminated which takes place in a relatively short time. This discovery led to the use of antitoxin, produced by actively immunizing horses, as a prophylactic to confer a passive immunity after wounds of a certain type are sustained.

This tetanus antitoxin was in use prior to the beginning of the last war but was not routinely administered to British troops until November 1914. It is interesting to note that the incidence of tetanus among the wounded treated in British hospitals dropped from thirty-two per thousand wounded in October 1914 to less than two per thousand in November and succeeding months. During the last war in our own Army every wounded man was given from 1,000 to 1,500 units of tetanus antitoxin as soon as possible after receiving a wound. Since antitoxin disappears from the body in about two weeks or a little less, in the case of contaminated wounds or fractures a second dose of antitoxin is often necessary in this method of treatment. This prophylactic use of antitoxin was very valuable in reducing the incidence of tetanus but was not entirely satisfactory since it could not always be given in time or in sufficient amounts to prevent tetanus and its administration was often followed by serum sickness.

The entire picture of prevention of tetanus was changed by the discovery in 1921 that formalin added to diphtheria toxin destroyed its toxicity without destroying the antigenicity. The resulting detoxified product was called anatoxin or toxoid. These results were confirmed in 1923 and the following year this method was applied to tetanus toxin and results secured similar to those previously worked out with diphtheria toxoid. Tetanus toxoid was first used by the French army as an adjunct to the use of tetanus antitoxin. When wounded under conditions that made tetanus probable, each soldier was given 1.5 cc. of tetanus toxoid subcutaneously in addition to a dose of antitoxin. This was followed in twenty days by a second dose of toxoid, and occasionally a third dose was given a month later.

The toxoid used first, and still in use to some extent, is made by growing tetanus bacilli in fluid media, from which the organisms are later filtered. To this toxic filtrate is added formalin and the resulting non-toxic preparation constitutes plain or fluid toxoid. In 1930 it was found that by precipitating this toxoid with alum a product of somewhat increased antigenicity resulted. This second product is called alum precipitated toxoid. Both products are in use at the present time; our Army uses the plain toxoid and the Navy the alum precipitated toxoid. Immunizations against tetanus in both Army and Navy are given as soon as possible after the individual's entry on duty.

The Army procedure is to give three subcutaneous doses of fluid or plain toxoid of 1 cc. each at intervals of three or four weeks. A routine stimulating dose is given at the end of one year in all cases. Prior to entry into the war zone all persons are given a stimulating dose. In the event of wounding or manipulation of old wounds

the individual is given a stimulating dose of toxoid instead of antitoxin as used formerly.

The Navy procedure is essentially the same except the alum-precipated preparation is employed. The initial immunization with the alum-precipitate consists of two injections of .5 cc. each, at an interval of not less than four nor more than eight weeks. The alum-precipitate is given intramuscularly. Stimulating doses are given as stated in the case of Army personnel.

Since this procedure was adopted by the Army early in 1941 a total of nine cases of tetanus have been reported for the entire army, seven of these had received no toxoid, the other two had had toxoid three to seven months prior to injury and no stimulating dose at time of injury. To date no cases of tetanus have been reported following battle injury.

THE DISCOVERY OF YELLOW FEVER VACCINE

The discovery of a vaccine for the prevention of yellow fever is one of the outstanding examples of the progress made in specific immunization procedures during the past few years. The control methods employed since the role of the Aedes mosquito was established have kept this country free of yellow fever since the first few years of this century and energetic control work in South America has greatly reduced the danger from that endemic focus. The increased use of the airplane and the exposure of our troops in endemic areas have to some extent offset the reduction of the danger which has been brought about by mosquito control measures. The vaccine can be relied upon to keep our troops free from infection and as a valuable adjunct to mosquito control should yellow fever be introduced into this country.

Yellow fever is one of the virus diseases and an attack of the disease gives a permanent immunity in practically all individuals. Every attempt to use a vaccine made from killed virus resulted in failure and attention has consequently been directed toward developing some method of using the live virus to provoke an active immunity. The early studies attempted to use the live virus partially neutralized by convalescent serum. The virus used in these studies was a strain which had been adapted to growth in the mouse brain and had lost much of its virulence but had retained its antigenicity. Mixed with human immune serum this vaccine was not adaptable to large scale production on account of the large amounts of immune serum required. It was then found that prolonged culture of the virus in embryonic tissue cultures prepared from mouse embryo, reduced the virulence while further propagation in the developing chick embryo brought about further reduction in virulence without destroying the antigenic properties. The strain of virus finally developed for use as vaccine is known as the 17 D strain. This strain was passed many times through rhesus monkeys and then established in mouse embryonic tissue cultures. After eighteen passages in this culture medium the strain was transferred to a medium containing whole chick embryo and after several passages in this medium was shifted to a similar medium from which the brain and spinal cord of the embryo had been removed. The virus has been maintained by subculture in this chick embryo medium containing minimal amounts of nervous tissue since its adaptation was secured. In vaccine production infectious material from this stock strain, either from the current culture or from material preserved in dried form, is used to inoculate living chick embryos about the seventh day of incubation. The inoculation is made through the sterilized egg shell and the eggs are then incubated for an additional three or four day period. After incubation the embryos are removed from the shell, pooled, finely ground, filtered, frozen, and dried in successive steps. After proper tests to insure safety and potency the vaccine is ready for use. To insure that the virus remains alive until use it is necessary that the preparation be kept dry and frozen continuously until it is administered. To use, the vaccine is rehydrated, diluted 1 : 10, and .5 cc. injected immediately. Originally human serum was added to the infected chick embryos but this practice was discontinued early in 1942.

This living virus vaccine was first tried on monkeys with resulting immunity and tests were then made on a small group of volunteers. No untoward reactions developed in these individuals and it was found that immune bodies could be demonstrated in the blood within a week to ten days after vaccination.

Slight symptoms and clinical signs usually develop in five to eight days following inoculations. Headache, slight fever, and "grippe-like" pains occur in some of the vaccinated individuals while no symptoms are noted by others. Large scale field tests of the vaccine were made in Brazil in 1937 and 1938 and immunity established in a high percentage of those vaccinated as indicated

by protection tests on serum specimens taken subsequent to vaccination. Recent reports state that only one recognized case of yellow fever has occurred among over 600,000 persons vaccinated with yellow fever vaccine in Colombia in the years 1937 to 1943, while 198 proven and 45 probable cases have occurred among unvaccinated individuals in the same areas during the same period of time. Since the vaccine has not been in wide use for more than six years the duration of immunity has not been established but data so far collected indicate that it will last at least four years in adults and probably longer.

At the outbreak of the war the policy was adopted of vaccinating our troops who were destined to travel to, or through, or to be stationed in areas where danger of yellow fever exists. While it is impossible to judge the number of exposures in our troops to this disease, it may be stated that no cases have developed in our armed forces since the adoption of this method of immunization to the present time.

THE DISCOVERY OF TYPHUS VACCINE

In order to clarify a discussion of immunization against typhus it is necessary to state that this disease is caused by rickettsiae which, like the viruses, can be grown only in the so-called tissue culture media containing living cells. There are two epidemiologic types of typhus, each of which is determined by the arthropod vector. One type called endemic or murine typhus has a reservoir in nature in the common rat. It is transmitted from rat to rat through the agency of rat fleas and rat lice, and occasionally to man by the rat fleas. This form of typhus is prevalent in our southern states and is gradually spreading northward. Murine typhus has a low case fatality rate, is sporadic in character, and there is no record to indicate that this form of typhus is communicable from person to person. The second type of typhus fever, usually referred to as epidemic, louse-borne, or classical typhus, is transmitted from person to person by the body louse. The relationship of these two types of typhus to each other is not known at present. The possibility that louse-borne typhus may develop from the flea-borne type has been suggested but exact knowledge on this point is lacking. Epidemic typhus has a high case fatality rate, 20 to 50 or 60 percent, is prone to cause epidemics, and is highly communicable from person to person through the body louse. This form of typhus is one of the great military diseases of history, having taken a severe toll of armies in past wars. Other severe epidemics have occurred in times of famine and social revolution.

Up to and including the last war the methods of controlling epidemic typhus fever were based entirely on the control of the body louse. Such control is not always simple in the case of hard-pressed troops, or disrupted civilians. Our methods of louse control have been greatly improved since the beginning of the present war and our method of immunization has been developed entirely since the last war.

The search for a method of immunization against typhus has followed two general lines; one has sought to develop a living virus vaccine by employing mild strains while the second line of investigation has been directed toward the development of a killed vaccine. The second method has proved to be the more satisfactory.

In the early attempts at production of a living vaccine tissues of infected animals were used in small amounts with the hope that only a mild attack of the disease would be produced. Also small amounts of infected tissues were mixed with the sera of immune individuals or animals that had recovered from an artificially induced attack of the disease. Results obtained from these methods were not at all satisfactory since neither the virulence of the living rickettsiae nor the degree of their neutralization by the immune serum could be definitely titrated. More recently other investigators, particularly the French, have utilized the infectious excreta of fleas infected with murine typhus as a source for the manufacture of a living vaccine. It has been found that the rickettsiae of typhus in the dried feces of infected fleas remain viable for many months when properly dried and kept under suitable storage conditions. Vaccines made from this source have been used extensively among natives of North Africa. From available reports the results from the use of this vaccine are uncertain in that the reactions produced vary from the development of definite and sometimes severe attacks of typhus to no reaction at all. In the latter event immunity seems also to be lacking. The reactions produced among Europeans are stated as being too severe to warrant its general use, although it is also stated that the reactions among natives in North Africa are milder. In any event, the use of living vaccines may be

summed up by stating that the vaccines of this type developed up to the present time are too uncertain for consideration as a means of immunizing our troops as long as an efficient killed vaccine is available.

In the early attempts to produce a killed typhus vaccine, the organs of typhus-infected animals were used as a source of rickettsiae and some chemical, or heat, was used as a killing agent. No great degree of success attended these attempts and the failure was generally and probably correctly attributed to the low concentration of the rickettsiae in the animal tissues. It was later noted by other investigators that the concentration of rickettsiae in infected lice was significantly higher than in animal tissues and a vaccine was made from the bodies of typhus-infected lice. Since human body lice only feed satisfactorily on human beings, this discovery remained of academic interest until a method of infecting lice by injecting them intrarectally with infectious material was developed. The infecting of lice by this method was developed by Weigl and the vaccine produced by utilization of this method is known as the Weigl vaccine. Briefly, lice infected by intrarectal injection of infectious material are fed on immune persons to allow multiplication of rickettsiae in the louse intestines. The intestines, which contain large numbers of rickettsiae, are dissected out and triturated in carbolized salt solution and constitute the vaccine. It requires the intestines of about one hundred lice to produce enough vaccine to vaccinate one individual. The Weigl vaccine was used, in the few years preceding the present war, as extensively as the supply permitted and apparently has been used by the Germans to some extent in the protection of doctors and nurses during this war. How extensively it is being used at the present time is not known but the tedious method of manufacture precludes its use on any very large scale. Reports of the use of this vaccine before the war indicate that it produces an immunity sufficient to protect against death and probably also reduces the number of cases.

The development and use of the Weigl vaccine showed that a definite immunity could be produced with a killed vaccine but the method could not be applied to mass immunization. It however encouraged search for a feasible method for growing rickettsiae. One method was developed in which the rickettsiae were harvested from scrapings of the peritoneum of typhus-infected rats. This method, like the Weigl method, did not lend itself to mass production. Two other methods were subsequently developed. In the first, the rickettsiae are grown in the yolk-sac of the developing chick embryo, while in the second, animals such as rats and rabbits are inoculated intranasally with infectious material. A rickettsial pneumonia develops in these animals in which the lung tissue is loaded with rickettsiae. In either method the rickettsiae are harvested and killed by the addition of formalin. The yolk-sac method, developed in this country, is preferable to the lung method from the point of view of ease of mass production and purity of the final product. Both preparations are used extensively at present, the lung type probably more extensively by the European nations than the yolk-sac vaccine, possibly on account of inadequate supplies of fertile eggs.

In the yolk-sac method fertile eggs are incubated for six or seven days. The infectious inoculum is then injected into the yolk-sac through a small hole in the air-sac end of the egg with a $1\frac{1}{4}$-inch 21-gauge needle. The eggs are incubated until death of the embryo which usually occurs in five to seven days. The yolk-sac is then removed, ground, and extraneous tissue removed. The antigen recovered in this process is suspended in saline containing formalin and one of the mercurials added as a preservative. After appropriate purity and potency tests the vaccine is ready for use.

This type of vaccine is under large scale production in this country and is used to immunize all of our troops destined for areas where epidemic typhus may be encountered. The vaccine is administered subcutaneously in doses of 1 cc. at weekly intervals until three doses are given. A recall or booster dose is given every four to six months or oftener while the danger of contact with typhus exists. From all available data it appears that this vaccine is as good or better than any of the other vaccines yet developed. The records indicate that protection against death is almost surely afforded and that the length of illness is shortened. It is probable that the incidence of the disease is also lowered although no satisfactory data on this point are as yet available.

SUMMARY

In briefly summarizing the recent most important immunological advances in war medicine it

may be stated that:

1. Typhoid vaccine has been improved in its immunizing potency.
2. Tetanus toxoid, yellow fever vaccine, and typhus fever vaccine have been discovered, improved, and their value established.

Further improvements in these and in other specific immunizing procedures will most certainly be made in the future. It is, however, comforting to know that the procedures just discussed are at present very satisfactory and that our troops are as well protected by specific immunizations as are any in the world.

HUMAN PROBLEMS IN MILITARY AVIATION

DETLEV W. BRONK

Professor of Biophysics and Director of the Johnson Research Foundation, University of Pennsylvania; Coordinator of Research, Office of the Air Surgeon, Headquarters Army Air Forces

(*Read April 20, 1944, in the Symposium on Wartime Advances in Medicine*)

ONE of the memorable experiences of this war is to stand at evening in the Great Court of Trinity College, Cambridge, outside Isaac Newton's rooms, and watch the Flying Forts return. Against the soft English sky they move in stately formation, seeming to proceed slowly because of their great height. Their silhouettes, floating by the hundreds through the wisps of clouds, move one with admiration for man's mechanical genius that has driven his machines into the skies, against the force of gravity.

A little later, in the gathering dusk the Fortresses roar down the runways of their scattered fields, and the machines are revealed as mere instruments of human crews. Waist gunners sit casually in their openings, waving as they pass; bombardiers are in their transparent cages; pilots taxi their ships to rest. They are men returning from a mission made doubly hazardous by enemy action and the dangers of an environment unnatural to man. These are the crews who make the majestic armadas of the air a symbol of man's new freedom from natural limitations, gained by courage and by science.

The history of aviation is a long record of man's restless urge to exceed the physiological restrictions of his body. It first appears in mythology and ancient literature as a desire for the birds' freedom from the gravitational tie to earth. Yet when in 1783 man was at last freed from his earth-bound life, he was but started on the conquest of his limitations. Each new power provided by the physical sciences has placed new stresses on the human body and has carried man into environments to which he has not been adapted by the slow process of evolution.

The first of these limiting conditions appeared soon after the pioneer balloon ascent. For by 1803 the science of aeronautics had so far progressed, that Robertson, the French physicist, succeeded in reaching an altitude of 22,000 feet, where he experienced an overpowering sense of indifference. "There," said he, "the scientist is no longer sensitive to the glory and passion of discoveries, and it is only by the aid of a little fortifying wine that he succeeds in finding intervals of mental clarity and power."

Driven by the urge for greater range of action, man's ingenuity devised means for ascent to ever higher altitudes, until finally in 1862 Glaisher, the English meteorologist, and Coxwell, his balloon engineer, reached the limits of human survival. They went to a reported altitude of 29,000 feet, but Glaisher was unconscious, and both would have perished had not Coxwell, paralyzed though he was, seized the valve cord in his teeth and released the gas by vigorously nodding his head. On recovering consciousness Glaisher voiced man's faith in the power of science to break the bounds of human limitations: "I certainly shall not take it upon myself," said he, "to set the limits of human activity and indicate the point, if it exists, where nature tells the aeronaut: 'You shall go no further.'"

His confidence was soon justified, for it was but a few years later that Paul Bert, the physiologist, discovered the cause of man's failure at high altitudes and prescribed the means for overcoming this restriction on human flight. Two causes were suspected: the reduced barometric pressure and the lower volume concentration of oxygen in the atmosphere. Both of these factors are now known to be a hazard to the airman, but the cerebral symptoms then described were shown by Bert to be the consequence of too little oxygen in the inspired air. This he proved by placing a man in a large metal chamber in which he could be subjected to one or the other of the two conditions.

From that time on, the vital dependence of the brain upon an adequate and continuous supply of oxygen has been well recognized. The recent development by my colleagues—Davies and Brink—of an electrochemical method for measuring the concentration of oxygen among the nerve cells of the intact cerebral cortex emphasizes this essential need. By exact measurements it is now possible to gain a clearer understanding of the

brain's requirement for oxygen and of the factors which regulate the supply of this essential element. Pertinent to this discussion is our observation that as the recording electrode is moved to points on the cortex farther and farther from a blood-vessel, the concentration of available oxygen decreases steeply until, at a point 25–50 microns from a vessel, the concentration may be less than a tenth of that in the arterial blood. This steep gradient of concentration insures effective diffusion of oxygen to the cells, but their avid consumption keeps the reserve supply at a low level. In consequence of this a deficiency of oxygen in the blood is quickly followed by a precipitate drop in the oxygen surrounding the cells. Rapid failure of nervous function then ensues.

To protect the cells of the brain against this danger the human organism possesses a remarkable system of reflexes. When the oxygen concentration in the blood flowing to the brain declines, certain nerve cells are stimulated to action, and that action causes an increase in the rate and depth of respiration. Thus more oxygen is again supplied to the blood flowing through the lungs. The heart rate is also accelerated, and more blood is sent to the brain in order to maintain the necessary diffusion of oxygen out to the nerve cells.

Even in the early days of flight such a sequence of events was observed to be a natural protection of an airman against one of the dangers of high altitude. For in 1804 Gay-Lussac reported that his heart-rate and respiration were progressively increased during the course of a balloon ascent. But as the aeronauts ascended to higher altitudes, where the oxygen was less plentiful, the physiological protective mechanisms were ultimately inadequate. To supplement these natural powers of adaptation it was, accordingly, necessary to provide some means for increasing the oxygen delivered to the lungs. To do this, Paul Bert simply provided small containers of oxygen, with tubes which could be inserted in the mouth.

For fifty years this simple expedient satisfied the needs of airmen, excepting the few who went to unusual heights. But now the strategy and tactics of aerial war make operations above 30,000 feet a routine event, and new problems have arisen. At those altitudes, the oxygen in the atmosphere is so low that care must be taken to seal the nose and mouth against the outside air, lest the oxygen drawn from the reservoirs be too much diluted. To accomplish this, facial masks have been designed which open to the air through a valve only during expiration. It is thus possible to deliver pure oxygen to the flyer.

If this be done for the ten-man crew of a bomber on a six-hour mission, hundreds of pounds of oxygen tanks are required. This, in turn, reduces the gasoline and bomb load by a corresponding amount, and the airmen, although physiologically protected, accomplish less on their military mission. To make the most effective compromise between human necessity and strategic efficiency it is, accordingly, desirable to supply just enough oxygen to satisfy the requirements of each individual under any condition of altitude or bodily activity—just that and no more.

The natural indicators of what this need is at any moment are the nerve cells of the respiratory centers which regulate the rate and depth of respiration. Their action can, in turn, be made to control the flow of oxygen to the aviator's mask, by placing in the supply line a regulating valve which is activated by the suction created by each inspiration. To gain further economy, oxygen from the tanks is mixed with outside air in order to utilize the oxygen available in the atmosphere. In order that this dilution be not too great at the higher altitudes, the mixing occurs through an aneroid-controlled valve. It will be obvious that the safe and economical design of such equipment depends upon precise physiological knowledge concerning the rate of oxygen supply required by the human organism, the individual variations in this requirement, the additional oxygen needed at any altitude, the tolerable inspiratory and expiratory pressures and the effects of varying degrees of work on the respiratory demands. Our modern oxygen supply systems are accordingly the products of close co-operation between the physiologist and the engineer. With this equipment men now go safely to 38,000 feet.

There new limitations appear, for the barometric pressure at those altitudes is so low that insufficient oxygen goes into the blood passing through the lungs, even though pure oxygen be delivered to the mask. Unless this situation is overcome, new aircraft designs and engine superchargers are not usable in outreaching the range of enemy anti-aircraft fire. To satisfy the human requirements at these great heights, oxygen must be delivered under a pressure sufficiently high to load the blood adequately.

Two methods are now available for accomplishing this. One is to design the oxygen regu-

lator, valves and mask so that a pressure can be maintained greater than that of the ambient air. The other is to seal the aircraft cabin and keep the interior at sufficient pressure by means of a mechanical compressor. The first is a temporary expedient. The latter is an ideal solution, for it gives the crew adequate oxygen with complete freedom of movement, unencumbered by mask and tubes leading to a reservoir. The sealed aircraft enclosure has the further considerable advantage of providing warm surroundings for men who must quickly go from a temperate or tropic climate to regions where the temperature is 40° F. below zero. With such pressurized cabins the engineers will have at last restored to fliers their natural environment, while taking them to altitudes unsuitable for life.

One of the most dangerous hazards of high altitude flight is the insidious character of the symptoms of beginning oxygen want. As the early balloonists described them they are lack of mental clarity or even a sense of well-being that masks the danger of imminent catastrophe. It is, therefore, not feasible to rely upon the judgment of an airman as to when he is in need of his supplementary oxygen supply, unless he be thoroughly familiar with his physiological requirements and the means for satisfying those demands.

In order to forestall such dangers, the Army Air Forces and the Navy have during this war instituted Altitude Training Programs for instructing every pilot and air-crewman in the basic physiological principles that relate to life at high altitudes. In the Army sixty modern replicas of Paul Bert's "altitude chamber" have been erected at forty-five air fields. Under the direction of some two hundred aviation physiologists, student fliers in groups of ten or twenty are placed in these chambers for several hours, while the atmosphere is made to resemble that which they will later find at 35,000 or 40,000 feet. Under careful supervision, they experience the effects of inadequate oxygen and are taught to use the equipment available in combat aircraft for providing the additional oxygen they will require. Finally, the "chamber flights" are supplemented with lectures which give the physiological basis for the practical demonstrations. Forty-five thousand men a month have been given such instruction, in what is probably the largest medical training program ever developed to overcome an occupational hazard. This is an essential part of the training that fits the crews to fly the great bombing missions over the Nazi's European fortress.

The problems I have thus far discussed differ only in degree from those encountered fifty years ago in slow-moving balloons. The high speed which gives to modern aircraft a unique tactical advantage now confronts the flight surgeon with quite new problems. For the aircrews not only ascend into an unnatural environment, but they do so with terrific speed. Climbing at a rate of forty feet a second, an airman in six minutes reaches an altitude of six miles, where the barometric pressure is but one-third that at sea level. This sudden change in the pressure acting on the body unbalances the equilibrium of gas pressures within its cavities and tissues. The painful outward pressure on the ear drum when the eustachian tube cannot be opened is a familiar experience of all who have flown. But only the military aviator going quickly to 30,000 or 40,000 feet knows the excruciating pain caused by the sudden liberation of gases from solution in the blood or other body fluids.

It has recently been shown that during these sudden changes of pressure minute gas nuclei on the surface of cells or on the inner walls of blood vessels rapidly expand in size, growing with the nitrogen, carbon-dioxide, and oxygen that is liberated from the surrounding fluid as it is decompressed. As bubbles of gas are thus formed and grow, some lodge in small terminal vessels where they obstruct the flow of blood. Nerve endings may thus be deprived of oxygen and pain results. Or regions of the brain are likewise put out of action, with widespread and serious consequences.

I must hasten to add that this widely accepted explanation of the cause of the pain of decompression sickness is not conclusively established. The theory has, however, served a useful purpose, for it has led to a practical and fairly reliable method of prevention. Assuming that the bubbles are largely composed of nitrogen—that being the most plentiful source of gas within the body fluids—it was suggested that the supply of nitrogen for bubble formation could be reduced by breathing pure oxygen for some time before a flight. Nitrogen in the tissues is thus replaced by oxygen. Because the oxygen is consumed by cellular metabolism, and because its tension in the blood rapidly falls when a little is removed, it is less potent as a source of bubbles. This has been proven experimentally, and the practice of breathing oxygen before a flight or on the climb

to altitude is now a proven means for preventing decompression pains.

Ultimately these symptoms will be unknown, for in the sealed-cabin aircraft of the future compressors will maintain the pressure constant throughout a flight to any altitude. In the meantime engineering science has given fliers machines which overtax their physiological powers of adjustment. To make possible the tactical employment of these aircraft in long-time, extremely high-altitude photographic reconnaissance or in rapid-climb bombing flights, the discovery of the preoxygenation procedure has been a great boon.

There is another notable instance in which the accomplishments of the engineers have exceeded the capabilities of the flier. Two of the primary requirements for good fighter-craft are high speed and great maneuverability. These are the characteristics which enable them to excel in plane-to-plane combat, to evade the heavier fire-power of larger craft, and to give effective protection to our bomber missions. Engineers and metallurgists worked for years to develop these planes that would withstand the centrifugal forces of high-speed turns, but during that time there were no corresponding improvements in the physiological mechanisms of the men who were to utilize these new machines.

A normal heart and circulation will deliver enough blood to the brain when the body is erect or recumbent, and will meet the needs during sudden changes in posture. Nerve messages, from pressure sensitive nerve endings in the walls of the carotid artery and in the arch of the aorta, promptly report to the nerve centers regulating the heart and blood vessels a drop in blood pressure in the vessels supplying the brain. The effect of this is an accelerated heart rate and a constriction of peripheral vessels. Thus the circulation of the brain is again increased. But the cardiovascular system and this reflex control were not evolved for pumping blood made five to ten times as heavy by a suddenly applied centrifugal force.

Under these conditions blood accumulates in the lower extremities and in the viscera. The result is an inadequate supply of oxygen to the nerve cells in the brain. "Grey-out," then, black-out" of vision are the first effects, but, if the centrifugal force be sufficiently great, and prolonged for some seconds, loss of consciousness will follow. In less severe maneuvers, there may be no obvious symptoms, but the often-repeated reduction of cerebral blood flow causes fatigue, irritability, and inefficiency. To help the human body meet these stresses imposed by the swift combat planes, belts and suits have been devised which aid the heart, by preventing the pooling of blood in the lower parts of the body. Without such devices the plane is but a futile object, unsuitable for use. For although a pilot may execute maneuvers that cause repeated black-out with no evidence of permanent damage, he is a relatively ineffective fighter under those circumstances.

Aircraft maneuvers which are quite inadequate to produce such cerebral effects exert forces on the human body which make unreliable the normal sensory mechanisms concerned with posture and position. The force which holds the plane in a banked turn or in a loop also acts upon the gravity receptors, the tension receptors in the muscles and the tendons, and the sensory mechanisms in the semicircular canals which detect rotation. These sensory pathways are then stimulated by the resultant of this machine-exerted force and that of the earth's gravitational field. Because the sensory mechanism is unable to resolve these two components of the stimulating force, a true sense of orientation in space is lost. The false sense of position must then be corrected by visual reference to the earth. And so, when clouds or darkness interfere, the pilot becomes incapable of maintaining a desired course relative to the surface of the earth. "The night is come; be then the night obeyed" was accordingly the rule of aerial combat in the First World War as it was for the foot-soldier in the Homeric battle between Hector and Ajax.

Finally, however, the causes of man's inaptitude for directed flight without visual contact with the earth were recognized. It was then possible to devise instrumental aids to the senses, and flight through clouds and darkness became safe and commonplace. Once again the physiological characteristics of the flier had limited his full utilization of the aircraft; once again medical scientists defined the need, and physicists provided instruments such as the artificial horizon, the bank-and-turn indicator and the gyroscopic compass to supplement the senses. And so the scope of human flight was once again increased. But, as we shall see, still more human problems were thus created.

The development of instruments that make possible flight by night has greatly extended the tactics and strategy of aerial war, for it enables

the airman to utilize nature's most effective form of camouflage. Unfortunately, however, this advantage is also available to the enemy, and this, then, requires a keen ability to see through the darkness the dim form of an enemy aircraft. Physicists are already at work extending the range of night vision through radar, but in the meantime men fly and fight with the aid of their natural vision.

To watch a night mission return to blacked-out Britain or to see hundreds of gliders sweeping into the unlit cornfields of North Carolina during an airborne maneuver, is to appreciate how unnatural such duties are to the boy who has grown up with an electric-light switch at his finger tip and a flash light in his pocket. Even the flight surgeon has had to recall forgotten facts concerning the mechanism of night vision: the fact that the cone cells of the retina which are used in day vision do not respond to dim lights; that the rod cells which are used in night vision are located in the peripheral regions of the retina, so that one sees a dim object best by looking a little away from it; that the sensitivity of the rods is destroyed for some time by a bright light, and is least affected by red of all the colors. The translation of these principles into tactical practice is an important element in the success of night operations.

A first step toward this end was a revision of lighting practice on airfields and in aircraft. Dark adaptation following prolonged exposure to bright light requires about half an hour, and until then night vision is below normal. Accordingly, a pilot who goes quickly from a lighted room into combat is relatively blind and under a severe handicap. He is likewise at a disadvantage on looking from his lighted instrument panel out into the darkness where enemy planes are hidden. To protect the airmen against these hazards, the Air Surgeon has recommended that lights to which the fliers are exposed shall be red in color, whenever that is possible. As has been said, this makes possible cone vision without affecting the subsequent sensitivity of rod night vision.

A more far-reaching effort to improve human efficiency in night combat is the establishment of a night visual training program. Following the pattern of the Altitude Training Program, the Army Air Forces are now developing a course of instruction which will familiarize the airmen with basic physiological principles relating to night vision; will warn him of the harmful visual effects of unnecessary light, of diets inadequate in vitamin A and of insufficient oxygen. To this is added several hours of directed practice in the detection and identification of faintly illuminated objects. We have reason to believe that it will thus be possible to improve visual performance in night operations by as much as 30 to 40 percent.

Just prior to our entrance into the war the National Research Council sponsored a survey of the sensitivity of rod vision in a group of several thousand military personnel. The results confirmed Hecht's laboratory findings that, within a normal population, some individuals are fifteen times better fitted by their retinal characteristics for night observation than the poorest, and five to ten times better than the average. It is accordingly possible to select with appropriate tests the individuals who are naturally qualified for exacting night duty, and to eliminate the unfit.

This is but an instance of one of the most important responsibilities of the flight surgeon in a technological war. In nature all men are not equal; the more exacting their tasks, the more important does it become to select those best fitted for specific duties. And so the medical services of the Air Forces have during the last thirty years developed searching tests to determine the physiological fitness of an individual to meet the stresses of aerial combat. But even these fail at times to reveal some subtle incapacity, and we have accordingly added to those tests certain realistic examinations under actual conditions of environmental stress. Such is the observation in an altitude chamber of each prospective airman for abnormal reactions to high altitude. If he reacts unusually to mild lack of oxygen, he may be rejected; for temporary emergencies arise with even the best oxygen supply equipment. Or an individual who is especially susceptible to pain during decompression is limited to low altitude operations.

Beyond these more obvious physiological characteristics of his constitution are the mental, emotional, and temperamental aptitudes of the candidate. The demands of constant attention to exacting operations, the frequent necessity for quick decisions of vital importance, and the constant threat of disaster place a severe nervous strain on an airman. Those who are considered to be incapable of withstanding these stresses are eliminated. But even those who must be judged mentally normal ultimately face severe threats to their endurance. To ease their strain, the medical examination of airmen has been broadened

to include a psychological assessment of a candidate's intellectual qualifications and his special aptitudes. Thus, the fitness of an individual for the highly technical work of aerial war is judged by an appraisal of the man as a whole—physically, mentally, and temperamentally. The chances of his success in his job depend upon the accuracy of this appraisal. What is even more important to the flight surgeon is that the chances for good mental health will be better if the man is given the work for which he is best fitted, and the danger of accidents or sickness will be less.

In a series of psychological research units each trainee is carefully assessed by interviews, written examinations, and psychomotor manipulative tests. Those who demonstrate good judgment, facility in learning eye-hand coordinations, and a high degree of emotional stability, with qualities of leadership, are preferred as pilots; accuracy of motor coordination in the execution of precise manipulations and a quick reaction time meet the requirements for a bombardier; an interest in mathematics and logical, precise habits of thinking suggest assignment as a navigator.

That the psychological examining stations are referred to as "research units" indicates the exploratory and experimental nature of this program. Criteria for acceptance into aircrew training and for assignment to specialized duties are frequently reviewed so that the predictions shall be as accurate as possible. To accomplish this the tests are constructed so as to meet the needs of the operations for which men are being selected. And the predictive accuracy of the tests is from time to time checked against subsequent performance, by assigning groups of men to duties without taking into consideration the psychological evaluation of their aptitudes.

The evidence gained from one large, typical group demonstrates the extraordinary value of this program of selection and classification. From among those of this group judged by the psychologists to be best qualified for the duties of a pilot only 4 percent were eliminated by the flight instructors during primary pilot training, whereas 78 percent of those who had been considered unsuitable by the psychologists were "washed-out" by the instructor. There was a similar accuracy of prediction for the subsequent performance of navigators and bombardiers. Also of medical significance is the value of this selection in the reduction of flying accidents. The 15 percent considered least qualified for pilot duty, on the basis of the psychological tests, had three times as many accidents during pilot training as had those cadets who had been judged to be the 30 percent best fitted for their duties.

I have dwelt at length on the human problems created by modern combat aviation. To these I might add certain others which arise from the wide-flung system of the Air Transport and Troop Carrier Commands. Quick movement from the Arctic to the Tropics subjects men to sudden changes of climate; forced landings in remote places taxes man's capacity to survive in strange and isolated environments; swift transportation of many troops favors the spread of geographically restricted diseases. But each problem has been met by the application of medical and physiological knowledge, and from this swift transportation has come a new advance in the practice of military medicine.

The same planes that carry weapons of destruction and personnel to the battle-fields bring back the sick and wounded, quickly and safely. Within a few hours after injury, a man may be hundreds of miles behind the lines in the security of a well-established hospital, where he is able to receive the best of modern medical and surgical treatment. The improvised surgery of the field station and the slow agony of hospital trains that tortured the body of the wounded and the patience of the tactician concerned with his supply lines may soon be only memories of the past. For in 1943 the Army Air Forces evacuated by air 173,527 sick and wounded patients (a figure that is not to be confused with announced battle casualties) with only 11 deaths in flight. This amazingly low death rate of 0.006 percent is effective refutation of those who doubted the medical feasibility of the Air Surgeon's far-sighted proposals in 1939 for the creation of air evacuation squadrons.

An important element in the success of this undertaking has been the employment of Air Transport cargo planes and Troop Carrier planes, rather than specially assigned aerial evacuation planes. By using the return flights of the cargo and transport aircraft, equipped with removable litters, planes have been conserved, and, what is more important, our aerial ambulances have been available wherever there is action. Once again aviation has reshaped the strategy of warfare and has given new advantages that reduce the sufferings of combat.

Few will be as bold now in predicting the future course of military aviation as was the founder

of the American Philosophical Society in 1783, when he watched Pilâtre du Rozier make the first balloon ascent. "This new invention may possibly give a new turn to human affairs," said Benjamin Franklin. For "it may convince sovereigns of the folly of wars, since it will be impracticable for the most potent of them to guard his dominions. Five thousand balloons capable of raising two men each could not cost more than five ships of the line; and where is the prince who can afford so to cover his country with troops for its defence as that ten thousand men descending from the clouds might not in many places do an infinite deal of mischief before a force could be brought together to repel them." Franklin's forecast may still prove correct, and the great aerial assaults of today may indeed "convince sovereigns of the folly of wars." If that comes to pass, it will be the greatest of aviation's notable contributions to human welfare.

ENVIRONMENTAL PROTECTION

GEORGES F. DORIOT

Colonel, Quartermaster Corps, Army Service Forces, War Department

(Read April 20, 1944, in the Symposium on Wartime Advances in Medicine, by Lt. Colonel D. B. Dill, in the absence of the author on war duties)

ENVIRONMENTAL protection, as it concerns the individual soldier, has been selected as a subject because of its increasing importance in fluid warfare. I am going to talk about protection from the weather, terrain, vegetation, and all those things over which the enemy has little control.

First I think I should tell you about the particular problems with which I am engaged—that is the research and development of Quartermaster items of equipment. These items may be categorized as things which a soldier requires in his everyday life and which contribute directly to his physical and mental well-being; or you may prefer to describe these as his purely defensive equipage. Our problem is to give him shelter, gear to sleep in, clothing to keep him warm or cool, protection from injury from inert objects or insects, food for his regular or emergency subsistence, materials for keeping him and his equipment clean.

I do not mean to give you the impression that the Research and Development Branch of the Quartermaster Corps bears the entire responsibility for the success of the American Army or its soldiers. On the other hand, I want to emphasize that without the development and supply of efficient personal equipage the soldier's fighting power is hampered. We like to look at food, clothing, and shelter as the soldier's first armor because his efficiency and productivity are dependent upon his physical state at the time of emergency; and not only his physical state; his mental outlook must be built up and it is surprising how much food and equipage may contribute to that end.

Significantly, the long arm of science has not touched such ordinary things as a pair of gloves or a buckle; inventive genius does not appear to have been encouraged, at least not with the same aggression as, for example, that applied to offensive implements.

I would like to give you many illustrations of the influence of personal equipment on the behavior of the soldier, but many of the examples must remain secret and the solutions to problems which we have found, and anticipate finding in the future, have to remain buried until the end of the war. At the conclusion of these remarks, however, I will be glad to answer such questions as I may in order to amplify those newspaper reports which you must have read, and to discuss ideas which you may have.

Protection from the environment concerns everyone. I want to talk particularly about clothing—foot, hand, and headgear—because everyone wears these articles and, deep down in our hearts, each of us is an expert on clothing. Who knows but that some modest expert in this audience will this afternoon interject the best idea yet received concerning protection from the environment. I recommend such an interjection.

Since you are all well aware of the history of clothing worn long before this war, I will concern myself principally with clothing in the last decade. Clothing has been largely developed for the purpose of dressing a retail store window and not for dressing the ultimate consumer. It seems that those sleek trousers, those wide padded shoulders and that smooth fitting drape, which some of us like to see but most hate to wear, have created for us a group of clothing and textile technologists who insist upon carrying similar techniques into functional clothing design. I do not mean to be critical of an industry. In the interests of promoting useful knowledge I hope I may be permitted to expose a number of misconceptions which this important industry holds.

If I tell you of our deficiencies in clothing design at the time the emergency occurred, an excuse is permissible because rarely has clothing been designed to suit a particular task. The functions and behavior of a soldier are not foreign to civilian practice. If you will stop to think of the particular tasks performed by, let us say, the construction worker operating a caterpillar tractor and sitting for hours on end in bitter cold, the miner kneeling on the damp ground, the telephone linesman performing dexterous tasks in icy

cold winds, the farmer tramping through mud, and even the civilian in our northern states trying to retain his balance on frozen streets, you will see that we really have not done much to provide them with the particular clothing and equipage which enables them to perform those tasks quickly, efficiently, and with a minimum of physical fatigue. People have said to me "Is speed, efficiency, and conservation of energy really so important?" In answering that question I come to the real reason why we devote so much attention to personal equipage. I cannot give you quantitative data but you must believe me when I tell you that casualties arising directly from enemy action form a relatively small percentage of the whole; the major part is made up by such things as bronchial disorders, malaria, foot ailments, frostbitten faces and fingers, shock, with the common cold well in the van.

We choose to regard this as a challenge; we regard it as a criticism of clothing, possibly of food, certainly of shelter and sleeping gear, the failure of which plays a tremendous part in reducing the soldier's physical resistance. I am not a physiologist, but I am assured that energy depends upon fuel, and extra fuel is needed to meet extra demands for energy as a result of environmental conditions. A high rate of cooling, such as results from wet sleeping gear and clothing, drains the soldier's stored heat. The conditions of operations do not permit the soldier to stoke up the furnace with unlimited food to replace that fuel at a time of high consumption, nor can he carry bulky protection or discard his clothing when he wants to do so. A good illustration of sapping the energy of a soldier unduly, may be found in the case of boots. In hot humid conditions of the jungle, every ounce of weight or, as we describe it "payload," demands a toll from the soldier when he is in only moderate activity. The result of adding weight is to increase sweat production, pulse rate, skin and internal temperature; the soldiers' tolerance to work and efficiency is progressively reduced the nearer the figures climb towards the point of heat exhaustion. A pair of combat leather boots weighs 5 lb., a pair of jungle boots weighs 3 lb. The addition of 2 lb. of weight to the footwear is equal in terms of heat stress to adding four times that weight or 8 lb. to the pack on the soldier's back. If any of you consider that you could walk as far, as quickly, and with your mental faculties as alert when carrying a 40 lb. instead of a 32 lb. pack, I recommend you try it in this pleasant climate. Imagine then, the strain which you would suffer at a temperature of 90° and 90 percent relative humidity. We find that the soldier in combat starts off with all the equipment, gadgets, and personal belongings he can pack; soon, things that you and I would consider essential, he discards. By bitter experience, he learns the price of weight, he supports and vindicates the physiological findings.

Weight is not the only drain on the soldier's energy or on his reserves of resistance. Bulk and volume impede his movements in the confined environment of a tank; on the ground they render him a more conspicuous target and impair his flexibility, his speed of action, his maneuverability. The soldier overheats in cold weather clothing because he may not be able to adjust closures, or discard parts of it when he is under high activity. Over 8 pounds of water may be added by sweat to the clothing over a period of time. Under the hot humid jungle conditions, clothing has to be worn largely to give protection from malarial mosquitos, from injury by vegetation and for camouflage—an unfortunate condition, for man would be much happier naked if humid heat were the only problem. Lightweight clothing wears thin and tears, and its resistance to the penetration of mosquito bites is soon lost. With more durable clothing, which is invariably heavy, a high physiological price must be paid by the wearer.

If boots are built tough enough to withstand the abrasive sand of the desert, an almost unbearable burning sensation of his feet results; if boots are light enough for evaporation or for squeezing water out, toes are stubbed, arches ache, and blisters occur; foot injuries are one of the surest and quickest means of becoming a casualty.

The soldier must stand for hours in inches of water in a foxhole or slog through mud. Boots which are good for wet conditions (and, unfortunately, there are none which are particularly good) are relatively useless for dry; but the soldier cannot carry two pairs of boots, nor yet many changes of socks. He sleeps in wet clothing. Evaporative cooling, compressibility of wet materials and their increased thermal conductivity increase the heat debt with a consequent lowering of resistance to cold and fatigue. Through his warm headgear, he may not be able to hear commands, movements of the enemy nor aircraft. His goggles, if they are dustproof, become fogged. If his handgear is thick enough for thermal pro-

tection, it ends up by being too bulky for the sensitive touch which fine instruments demand. Not the least problem is that of combining combat qualities in clothing with those of a military appearance; morale and a higher discipline reap benefits from such a compromise.

One of the greatest difficulties in providing equipment for the soldier is to predict his activities, because in reality his activities pre-determine the range of usefulness of his equipment and the minimum thermal protection which he requires. Our flyers have similar problems, although not, I think, as unpredictable, because the performance of the air craft assures you that certain minimum temperatures for a certain fixed period will be encountered. This does not happen to the foot soldier; he may be in a foxhole for three days, sometimes in six inches of water, or he may be in a particularly dry cave; his activity varies unpredictably from high to low; we cannot tell.

The problems are unlimited and unfortunately the ideal protection for any particular condition is rarely anywhere near the most efficient for the average condition; and so, one must compromise. I do not like compromises, I refuse to believe that a pair of combat boots cannot be reduced in weight from 5 lb. to 3 lb., and yet give equal efficiency, durability, and protection; I refuse to believe that closures whether at the ankle, the waist, the neck, or elsewhere cannot be made as stormproof as the textile or leather from which the clothing and footgear are made. Until a short time ago I was told that waterproof fabrics could not be produced without interfering with the wearer's sweat mechanisms and consequent comfort. I am happy to say that industry has achieved extraordinary results in this regard. I refused to believe that wool had to shrink, that a soldier had to wear tight socks (and tight socks are cold socks), until the industry dug its teeth into this problem and is now on the way to minimizing wool shrinkage.

I refuse to believe that the textile and clothing specialists, guided by physicists, chemists, and physiologists, cannot more closely approach all-purpose articles to satisfy diversified military requirement.

An illustration of the wide extremes of environmental conditions encountered will provide you with quantitative evidence of the need for protection.

Thermal balance is important under all conditions. In order to define simply the thermal protection afforded by the environment whether it is that of clothing or of the air within and around it, we use a unit named "clo." One clo is equivalent to the protection given by an ordinary suit of clothes which permits a sitting man to rest in still air at 70° F., and yet maintain thermal equilibrium: or, one clo is equivalent to the thermal insulation of $\frac{1}{4}$ inch of still air or of enough textile material to immobilize that air, if you care to interpret it by physical measurement.

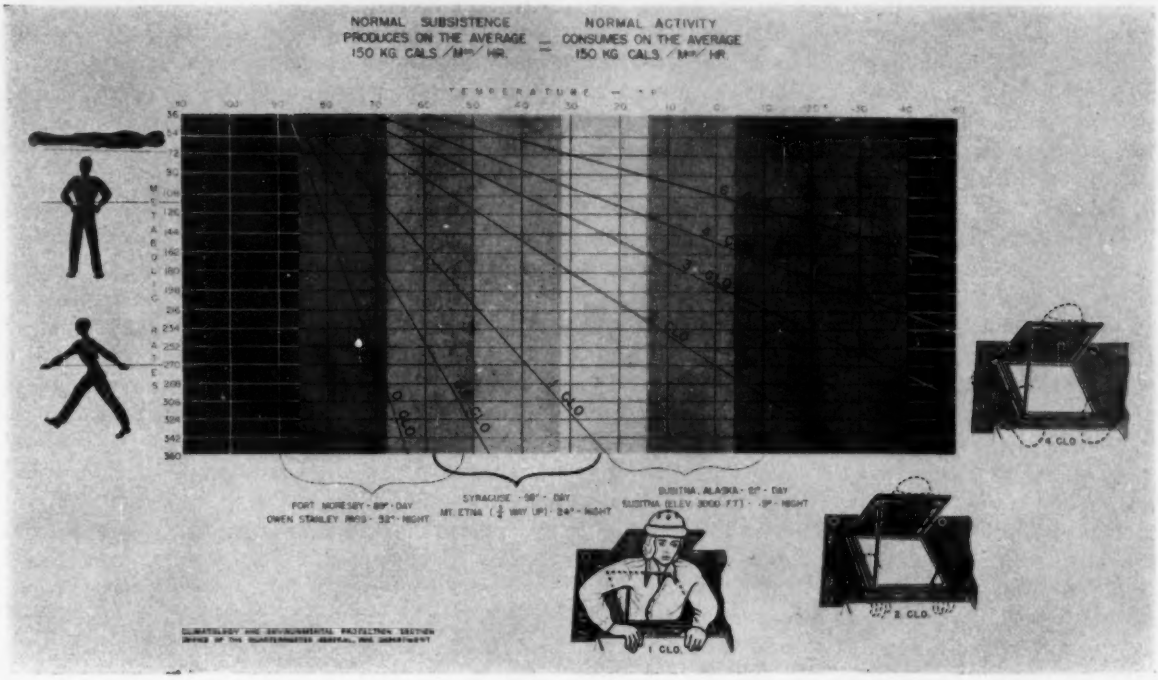
Approximately 4 clo is the maximum thickness which a soldier can wear and yet work in, and 8 clo the maximum volume he can carry as sleeping gear; above 2 clo on the hands does not permit the dexterity required. In still more realistic terms, a man who is deficient 1 clo in an environment which demands the maximum thermal protection is losing heat or energy, whichever you like to call it, at a rate 25 percent faster than he can produce it when he is inactive.

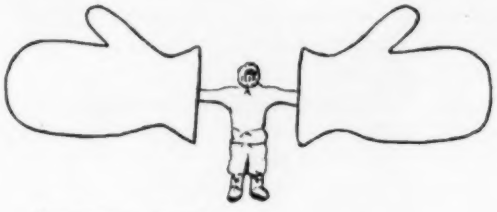
A man whose clothing is just right for a given temperature and for a given rate of activity will lose heat as fast as he produces it and no faster. If he is sleeping, he generates about 70 kilogram calories per hour; while standing guard he generates about 120 kilogram calories per hour; if he is walking, 270 kilogram calories per hour; if he is carrying a pack, he generates 360. Unfortunately we have not been able to put any one outfit on a man, which will give the right insulation at every level of activity or at all temperatures.

But in war you have to try to do this seemingly impossible thing, because men cannot carry a lot of changes with them, especially in battle. You have to make the best compromise you can and to make that compromise wisely in any particular case; you have to know quantitatively just what the clothing requirements of each particular situation are, and how the situation is likely to vary throughout a day. Here, this diagram may prove helpful. It is based on many experiments in physiological laboratories.

It shows various levels of activity, plotted vertically, from sleeping at 70 kilogram calories per hour to marching with a pack at 300 kilogram calories per hour. Horizontally it shows temperatures with color bands of climatic zones. The diagonal lines show the amounts of clothing converted to clo units which are necessary for thermal equilibrium, that is, the amount a man must have on for any given temperature and work rate, if he is to remain comfortable and not grow hotter.

Environmental protection over a period of 24 hours


NORMAL SUBSISTENCE
PRODUCES ON THE AVERAGE
150 KG CALS / M² / HR
=
NORMAL ACTIVITY
CONSUMES ON THE AVERAGE
150 KG CALS / M² / HR
TEMPERATURE — °F
4 CLO
3 CLO
2 CLO
1 CLO
0 CLO
FORT MORESBY - 89° - DAY
OWEN STANLEY PASS - 52° - NIGHT
SYRACUSE - 58° - DAY
MT. ETNA (½ WAY UP) - 24° - NIGHT
SUSITNA, ALASKA - 0° - DAY
SUSITNA (ELEV. 3000 FT.) - -9° - NIGHT



Any exposure, at rest

6 hour exposure, at rest

Best possible mitten, good for 2-3 hours, at rest

Strenuous exercise, no mitten needed

FIG. 1. Relative size of mittens needed for different exposure times at −20° F.

Thus, 1 clo, the amount you and I have on now, is enough to keep a man in thermal equilibrium for light activity at 70° F.; but if he lies down and sleeps, his metabolic rate is greatly reduced; he will want more insulation or a higher temperature around him; and if he works hard physically, he will want a lower temperature or he must discard certain parts of his clothing. Now let us see what we can do with this chart in a practical case. I take the case of a soldier marching in the environment of Syracuse in Sicily. He is wearing clothing equivalent to 1 clo protection because he knows that he is proceeding to a colder environment. According to the chart, he requires only ½ clo which means that his metabolic output is about 100 kilogram calories in excess of requirements; let us call it wastage of energy. He moves to a position at about 7,500 feet up Mount Etna. At midnight, he is standing on patrol in a slit trench; the temperature has dropped some 35° F., but he has been unable to carry along additional clothing and his task does not permit any compensating activity; under such conditions, he would normally require slightly over 3½ clo thermal protection. A drop of 35° in temperature and a change of activity from marching to standing, results in a net deficiency or energy expenditure of 250 kilogram calories; this is actually a heat debt of some 50 percent in excess of production. If the soldier had wanted to sleep, he would have required a sleeping bag providing about 6 clo protection; but this bulky item could not be carried, because everything but weapons and food would have been left behind. Without a bag and endeavoring to sleep, he would lose heat at the rate of about 300 kilogram calories per hour and this would mean depleting his energy reserve at exactly that rate.

There are three things this soldier could do; first, move about, exercise, and push his metabolic rate up; this robs him of rest and besides, in combat, he cannot get out of his hole and expose himself. The next thing he could do is to shiver. Shivering constitutes exercise and pushes up the metabolic rate but again he is robbed of rest. The third thing he could do is to do precisely nothing about it. If this process goes on long enough he will first of all become ineffective and shortly after become a casualty, either from illness brought on by exposure or because his body temperature goes down so far that he dies. You can die of cold at temperatures far above freezing; all that is necessary is to bring the temperature of the vital organs down a little below 80° F.

The soldier is not even comfortable at rest in humid heat. In marching he depends on evaporative cooling for maintaining heat balance. The clothes he wears for protection against mosquitoes and thorns add to the difficulties of temperature regulation. Unfortunately I know of no device for the soldier to wear which is cooler than no clothing; his maximum rate of marching consequently is reduced considerably; for at the normal rate, heat exhaustion would result very soon. This particular soldier will encounter a midnight temperature high on the Owen Stanley mountains of 52° F. where 3½ clo is required for his full sleeping protection. By adding the weight of such protection to the pack on his back, the daytime metabolic rate is raised even further.

Two further illustrations show the impracticability of providing adequate protection. Note on the chart the illustration of a man escaping from the hatch of a tank whilst wearing the bulk and thickness equal to 1 clo clothing. If bulk and thickness are increased enough to provide him with 2 or 3 clo thermal protection, he is either unable to escape from the tank or, if he escapes, he may have to leave behind him the clothing he wants for outside.

This other illustration shows the real size and bulk of adequate thermal protection for the hands at a wide range of ambient conditions. No explanation is necessary of the apparently hopeless task of providing hand protection in which dexterous tasks can be performed by an inactive man at temperatures of −20° F.

What is our role? It is to develop flexible clothing either warm or cool as need requires. There are several things we can do and many more may be possible, a challenge to our ingenuity for which no military tactics in the world can substitute. Scientific research in textiles and clothing must be relied upon to develop clothing which is adaptable to immediate protection from a condition of high activity on warm days to low activity on cold wet days with an infinitesimal increase in bulk and weight.

The failure of our personal apparel in peacetime has not been considered very important because if it rains, we can take shelter; if it is too slippery to walk, we can stay at home; if our hands are cold, we can put them in our pockets and even if we do not put them in our pockets, the compensating mechanisms of the body take

care of our abuse. The combat soldier on the other hand has no such leeway and his efficiency must be kept at peak for all conditions. Why is it not desirable nor economical for occupational specialists to be provided with the apparel which permits peak efficiency in peacetime? Why indeed should not all industrial workers be occupational specialists in the real sense?

We have learned many lessons from this war but I think the outstanding lesson is the absence of an almost elementary technical knowledge among the majority of personnel engaged in certain industries whether during war or peace. Loose thinking predominates in the claims made for certain raw and converted materials; but a little study of and thorough acquaintance with a few subjects is the type of knowledge which is really useful. There is a wide gap between research workers and those who should normally be called upon to apply the results of research. As a consequence, certain industries ignore what science has to offer.

I would like to discuss an anomaly in our education picture. When a student graduates as a lawyer, a doctor, or an engineer, he becomes a professional man. He deserves the rank he then attains, for he has studied and he has become adept in a specialized field. The backlog of knowledge arising from such concentrated study prepares him for all variables of his future professional life; it makes him receptive to innovations and it gives him the spirit to meet a particular situation with the background of at least theoretical experience. Other professional men in many industries are entitled to the same recognition and get it. I do not think many of those associated with the textile and clothing industries have been educated with the same intensity. I count among my very close friends leaders in textile and leather technology, millmen and clothing designers who have performed wonders for the soldier. On the other hand, when I raise a certain question, discuss a radical approach, I get the reply that there is no answer to that particular problem and that a fundamental research study will have to be initiated. I do not know many clothing designers who can recall, from their background in the industry, its patents, its leaders in certain specialties, and its technical accomplishments, as the lawyer would remember the case of Greene vs. Greene in 1906. In textiles, while the synthetic industry has approached its problems in a highly scientific manner, the individuals in that industry with very few exceptions, do not have the background or fundamental knowledge which could have saved all of us many months of searching, and the duplication of one another's efforts. The achievements of other industries or of other sections of the same industry have not been known and obvious solutions have been ignored. I wonder very much if those concerned with the raw textile components of fiber and cloth are as closely associated as they should be with the converted product and its purpose. Remember what I said about the miner kneeling. Our soldiers may shortly kneel with less stresses when they have similar tasks to perform because our trousers are designed for more than standing up in, and our gloves may be designed for only the clenched hands. Nobody, however, has yet invented for us a closure at the neck which will keep out the cold, the rain, and the snow, and nobody has yet shown us how a boot may be put on in a fraction of a second without cumbersome lacing. The resilient properties of down in sleeping bags have not yet been duplicated and I believe that those concerned with down substitutes never worried very much that a soldier must carry a bag which will keep him warm at 20° below zero and which must weigh only 5 lbs.; that it must retain its thickness for sleeping in after it has been washed. I do believe that a sleeping bag of this type even from a purely economic viewpoint is required in civilian life as an emergency item for air passengers, for trappers, for hunters, for construction gangs. Do you not agree that it would be desirable in civilian life to find a way of closing your collar satisfactorily (if a collar you must have) so that it keeps you as protected from the elements as does the rest of your clothes? I suggest that the textile and clothing industries encourage young men to embark early upon a career of training, and most important, to specialize in a particular phase, whether it be yarn, dyeing, clothing design, zippers, collars, or looms; that the training be with the aim of pursuing their trade with the same diligence as the lawyer pursues his; that we come to realize that the textile and clothing industries do not provide occupations to and from which one may drift and yet expect to be successful or contribute much to the welfare of others; that the clothing industry should be looked upon as having an unlimited scope ahead of it in designing, and using that functional design, if you like, as a merchandising weapon, with the material and cut features built to fit the wearer's particular task. I do not believe, and I hope you

agree with me, that we can possibly create a nucleus of qualified personnel unless we take advantage of scientific study, research, development, and testing.

Let me illustrate a point where, with sufficient curiosity, the textile and clothing trade would have called science to its aid with probable advantage. How many times have we heard the expression "Everybody knows that you can't wear impermeable clothing." The word porous is often used with very little further thought as to what kind of impermeability we are so loosely discussing.

If you care to picture the mud and cold wet conditions of modern warfare, you would be more inclined to re-word that common claim to"Everybody knows that you can't wear impermeable clothing with comfort but at least you may be a little more comfortable than being wet through in permeable types of clothing."

If you stop to analyze the discomfort of impermeable clothing, that is, textiles which are impermeable to free water and water vapor, that discomfort will be found to be primarily associated with body cooling. Allegedly, the body suffers when the escape of water is prevented, but I do not believe proof exists, at least in temperate climates.

Science on the other hand, could have discovered the effect of not wearing impermeable clothing in cold, wet conditions; it might have found that if a quart of water or rain enters the clothing from outside, 15 kilogram calories are required to raise the temperature of water at 41° F., to 68° F.; this heat so wasted on heating the water is equivalent to one fifth of the heat produced in an hour by a resting man. This is a wastage of 20 percent. This however is but a small wastage; as soon as the absorbed water in the clothing becomes warmer than air, it begins to evaporate. To evaporate that quart of water completely requires about 500 large calories which is over six times the heat produced by a soldier during an hour's rest.

Thus it may be understood that vapor-permeable clothing stacks up a large heat debt when wet; clothing which is perfectly adequate at zero when dry, will be hopelessly inadequate when wet at plus 50° F., if worn for but a short length of time.

If science had pursued this investigation further, it would have found that the reasons for the discomfort of vapor impermeable clothing—your common waterproof garments—are that normal clothing underneath it becomes slowly wet through by perspiration and that this moisture gradually starts a vicious cycle of vaporizing and moving toward the exterior where it condenses on the cold external waterproof layer and gives up its precious load of heat to the exterior only to soak back into the clothing for another load of heat. As more water is added by the body to this distillation system, the speed and intensity of cooling increases. Removal of the outer waterproof barrier and the increase of air movement greatly accentuates this evaporative cooling which may have a cooling effect equivalent to an air temperature drop in calm air temperature of 40° to 50° F. Moreover, the thermal conductivity of water is high compared to air, so that the insulating properties of air-laden textiles are greatly reduced and they become more or less a thin plastic mass under such conditions.

With all these facts before it, textile and clothing research workers might have considered that a scientific adaptation of the good points of impermeable and permeable clothing would be made; they might have found that a vapor barrier near to the heat source, i. e., the body, and another on the outside of the garment between which could be sandwiched a permeable thermal insulating layer, would achieve the desired end. By placing one vapor barrier near to the heat source, the vapor pressure thereat would be somewhat near to that at the skin with a consequent slight gradient such that evaporative cooling would be reduced tremendously. Air movement would be practically non-existent between the impermeable barriers.

There are many army tasks which do not call for high activity and high sweat production. I do not know if vapor impermeable clothing can be used, but we have certainly been wrong in arbitrarily saying it cannot.

Our technologists in textile matters have inclined to the thought that tensile strength and tear strength predicate the ultimate durability, comfort or the usefulness of a textile or garment. The Army has had the opportunity of setting up field tests with a large number of samples and from it ascertained a certain pattern of wear which is associated with some other physical property but doubtfully represented by tensile and tear strength. It has been the Army's fortune to show that the physical tests on which the industry has placed its confidence for years and years are probably unreliable and certainly not

representative. With vision and a little more realism, I am sure that the industry could profitably have diverted its powerful scientific ability to analyzing the durability, functional characteristics, and performance required by the consumer and, thereafter, it could have devised the physical tests which duplicated the actual conditions.

I feel sure that the rather bitter lessons learned by the Army will be valuable to industry after this war. I would like to close this address by urging this Society to foster facilities for thorough technological, together with liberal, training so that the quest for knowledge by the humble loom fixer, spinner, or seamstress may be satisfied. From these trained young men and women, given opportunities for advanced professional education and research experience, will come leaders in the clothing and textile industries. These professional experts, call them textile engineers, will provide the Quartermaster with facts rather than opinion as a background for action if and when another war comes.

NEUROPSYCHIATRY

EDWARD A. STRECKER

Professor of Psychiatry, School of Medicine, University of Pennsylvania; Consultant to the Surgeon-General of the Navy, and Consultant to the Surgeon-General of the Army and to the A.A.F. for the Secretary of War

(Read April 20, 1944, in the Symposium on Wartime Advances in Medicine, by Lt. Commander Joseph Hughes, Medical Corps, U.S.N.R., in the absence of the author on war duties)

In incidence, neuropsychiatric disorders and casualties are conceded the doubtful distinction of being the most important medical issue of modern war. Economically estimated at the World War I cost of $34,000 per patient, the financial forecast of the neuropsychiatric bill for this global conflict would soar to astronomical figures beyond the fantasies of even a psychiatrist.

Psychiatrists are more concerned about the relationship between neuropsychiatric disabilities and the morale of troops. In training areas an extremely large segment of the available hospital bed space is needed for these patients; their care and treatment, and often the administrative work in discharging some of them consumes a large amount of the available time of the relatively small number of psychiatric medical officers, and of course, while not through their own fault, it is still true that from these men who are sick in their minds and spirits there radiates ever widening circles of lowered morale. On the field of battle, a physical wound, even though it be mortal, appreciably raises the morale of those soldiers who witness it. They have seen a comrade, often a friend, stricken in his tracks. Under the triple lash of rage, hatred of the enemy, and desire for revenge they become, what in war is prized above all things—efficient, death-dealing, fighting machines. On the other hand, an hysterical convulsion or an uncontrolled exhibition of fear is strange, mysterious, and even frightening, and dangerously lowers morale.

There is a formula which is explanatory of all the neuroses of war, whatever be their particular clinical types. It may be expressed in terms of a universal and unconscious underlying emotional conflict, between man's most ancient endowment which he shares with all living things, his instinct of self-preservation and "soldierly ideals," a group of reactions acquired in the process of training and discipline in becoming a soldier. The first dictates caution about placing life in jeopardy. The second motivates behavior, which of necessity places life in danger. The first is very ancient and very strong; the second is more recently acquired and comparatively weak. In any event, in a war neurosis, the conflict is expeditiously, pathologically, and unconsciously resolved by the appearance of certain symptoms like hysterical blindness; deafness or paralyses; catastrophic nightmares; "startle" reactions or other manifestations of anxiety, which preclude return to the zone of danger.

In the face of threat to life from reality, even animals employ their tried and trusted weapons. Animals like elephants or rhinocerii utilize their tremendous poundage and meet danger by charging attempts to annihilate it. The large cats, like lions and tigers, meet their dangers by cunning, and the biting and tearing power of fang and claw. The oppossum feigns death. Myriads of insects achieve inconspicuousness and save their lives by merging their delicate camouflage into the trunk of a tree or a blade of grass. Likewise do human beings, when their psychological security is threatened, utilize the psychological weapons best adapted to their individual personalities, perhaps regression, rationalization, projection, repression, and other mechanisms. Thus, it is likely that the particular kind of war neurosis, conversion hysteria, anxiety states, neurasthenia, obsessive-compulsive reactions, is in large manner clinically determined by the distinctive markings of the personalities of the patients.

I do not believe that war in itself will produce new neuropsychiatric entities. However, it is unquestionably true that in this war as never before there has been a very heavy impress of extraneous factors. Such factors as fatigue, exhaustion, tropical diseases, and catastrophic emotional experiences may so shape the clinical pattern, that at the first casualty station it may be impossible to distinguish psychosis from psychoneurosis. Something depends upon the length of time the men have been exposed to physically and emotionally devastating conditions. On the carrier "Wasp" there was little sign of physical

stress or emotional trial until she was struck by three Japanese torpedoes and sunk forthwith. There were only two minor psychiatric casualties and the morale of the crew was superb—a naval epic. On the other hand, on Guadalcanal, where day after day, week in and week out, in sodden jungle, the men were subjected to indescribable hardships, privation, exhaustion, and emotional horror, the neuropsychiatric casualty rate was relatively high.

Conversion hysteria is a psychoneurosis in which, unconsciously and somewhat naively, the underlying emotional conflicts are converted into crude clinical signs and symptoms, often total annihilation of functions: deafness, blindness, paralyses, anaesthesias, aphonia, and the like. Conversion hysteria, the "shell-shock" of the newspapers, was the common neurosis of World War I. Up to the present time in this conflict, it is relatively less common and the large segment it filled is now occupied by anxiety reactions. There are several possible explanations for this shift in incidence:

(1) Thus far only a comparatively limited amount of infantry has been engaged and conversion hysteria in large preponderance is the psychoneurosis of the "run of the mill" man in the ranks and not of the superior enlisted man or the officer or of the flying personnel.

(2) In a brief span of less than twenty-five years, technical science has succeeded in producing not only much more lethal but far more emotionally devastating machines of war than ever before in our long and bloody history. This war is not a war of trenches and dugouts but of shallow fox-holes, and it is being fought in dank jungle and on burning desert sands; in subarctic wastes; in the air and beneath the surface of the sea. Furthermore, it is possible that deeper protective layers are being stripped from the emotional core. At least, the prevalence of anxiety in the war neuroses would seem to indicate that more profound emotional levels have been tapped. Such reactions as the catastrophic nightmare with the portrayal of stark fear and the "startle" reaction in which generalized and severe bodily tremors are elicited by unexpected noises are quite common. Mira, Psychiatrist-in-Chief to the Loyalist Spanish Army, described progressive fear profiles from mild motor over-activity to deep stupor.

Some of the prognostic indices largely derived from World War I have become axiomatic in military neuropsychiatry.

I. The shorter the time elapsing between the occurrence of the casualty and the first psychiatric treatment, the better the prognosis.

II. Within reasonable limits the closer to the line of combat the treatment is instituted, the more likely it is that the outcome will be favorable.

III. The soundness and integration of the premilitary personality or the reverse is reflected either favorably or unfavorably in the eventual result.

IV. The more severe the impact of battle, with exhaustion and harrowing emotional experiences, the more hopeful the prognosis.

V. The outlook is relatively better in anxiety reactions and conversion hysteria than in neurasthenia and obsessive-compulsive reactions.

TREATMENT

I have recently returned from an "overseas" mission, and therefore I am able to report on the methods of treatment which have been utilized in the various theatres of the war and which will be given at the front during the anticipated invasion of the continent.

One must not expect elaborate treatment facilities and techniques at the front. There is neither sufficient time nor psychiatric personnel available. These disadvantages are more than compensated for by the acuteness of the conditions and the rapidity with which treatment is instituted so that the results are more satisfactory than in civilian practice.

The chief reliance is on long familiar and simple therapy—rest; plenty of good and hot food; removal of symptoms by suggestion; reassurance and desensitization of the ego from the insult of not having been able to continue in action; the development of insight by an explanation of the nature of the underlying conflict and the mechanisms involved in the production of the symptoms.

This war has stimulated the modification and useful application of certain methods used in civilian psychiatry, notably narcosis therapy. In the method used in the North African campaign, narcosynthesis, a state of twilight consciousness is induced by the administration of sodium pentothal. When a state of mental confusion with removal of inhibitions has been attained, the patient is stimulated into verbal and motor activity by presenting to him his battle experiences. Often, it is obvious from the verbal and motor exhibitions of stark fear that the soldier

is not only remembering but is actually reliving these disrupting battle experiences.

At the front, the combination of rest, food, simple psychotherapy, and brief narcosis by the administration of sodium amytal is the therapeutic choice.

In the general and station hospitals in the "overseas" theatre a combination of mild narcosis and insulin is proving effective. The narcosis with sodium amytal insures the needed rest to the psyche and the insulin produces an average gain of weight of twelve pounds in two weeks, which is beneficial both physiologically and psychologically.

In base hospitals "overseas," far removed from the combat sector or in the continental limits, in addition to the usual psychotherapeutic procedures—occupational therapy and a carefully co-ordinated routine of daily activities—the most effective therapy is group psychotherapy. The patients discuss their problems together and the discussion is directed into useful channels by the psychiatrist group leader. Often, it is important to secure an emotional outpouring from the patient. The bar of inhibition may be broken in various ways: group discussion may soften resistance and hostility; so-called psychodrama in which the patients re-enact their experiences in a play; mild narcosis and resort to individual conferences with the psychiatrist are quite valuable in breaking the dam of the patients' inability and unwillingness to talk about their devastating experiences.

PREVENTION

The tremendous increase in the incidence of war neuroses and other neuropsychiatric disabilities in this war as compared to World War I has, in effect, served notice to the nation that more careful screening must be practiced and, furthermore, that in training areas soldiers must be thoroughly conditioned not only against physical hardship, but against the psychological hazards of war.

There are many difficulties in the way of thorough and completely effective psychiatric examination at induction. There is a serious dearth of available psychiatrists and since the stream of inductees passing through an induction station must not be "bottle-necked," not more than five minutes and usually less is all the time that can be given for the psychiatric phase of the examination. The majority of psychiatric disabilities, unlike many physical ones, deformities, spinal curvature, hernia, and the like, are not so obvious that he who runs may read. A large segment of psychiatric diagnosis depends on the history of the individual. Up to the present time it has not been feasible to secure histories in time for the induction examination, but a plan is now in operation which it is hoped will overcome much of this handicap. We all fervently hope that never again will we have to face a war crisis, but should this calamity occur, we would be better prepared if there were on record a survey of the national health, and if the war service act were less selective and mobilized every citizen from eighteen to seventy, so that each one could take his or her proper place in the total war effort.

MORALE AND PREVENTION

No one can doubt the intimate relationship existing between the size of the military neuropsychiatric problem and morale. Any consideration of prevention which failed to stress this relationship would be inadequate.

Great military leaders have always understood the tremendous significance of morale. Napoleon asserted that only one-fourth of victory was in the number of men and amount of material; three-fourths was in the spirit of the armies. He knew that while an army marched to its objective on its belly, it took the objective by its morale.

What is morale? Obviously it is more than the chemistry of man; more than the aggregate of his organs and parts or their functioning; more than the total of his mental functions. Man is much more than the sum of his parts. Morale is too intangible, too real, and too significant to be confined within a framework of words. Perhaps it is faith and courage and, in a national sense, devotion to the nation; desire to live for it and, should the need arise, the willingness to die for it.

Morale cannot be counted on to arise spontaneously. In the soldier, as in any one, it must be produced, honestly but deliberately.

The foundations of morale consist of simple, even obvious things—reasonably satisfactory living conditions and good, appetizing, well-cooked food; comfortable, nice-looking uniforms and, above all, easy fitting shoes. There should be interesting diversion and sports.

Upon this basic foundation, the superimposed layers of morale making ingredients contain larger and larger segments of psychological factors.

The medical care furnished the troops must be superior in quality and so given that the soldiers have complete confidence in the medical officers and the treatment, not only for the current illness but for any sickness or emergency that might arise in the future, either in encampment or in battle.

Soldiers should have a proper relationship with their officers—commissioned and non-commissioned. This relationship cannot be defined in all its ramifications. The contact should be such that the soldier will not hesitate to talk over and find counsel about his troubles and problems—not only military ones but home difficulties. Too many soldiers have had their morale broken by homesickness or by a complaining or sad letter from home.

Mass exercises and military drills have definite morale-making value. From such unison, mass activities, the confidence and security that come from the strength and bond of numbers are derived.

The condition of men for campaign and battle is highly important, not only for the promotion of physical endurance, but also psychologically. It is necessary to subject the soldier in training by day and by night, at least in miniature, to the sudden and surprising noises and the horrible sights of war.

No soldier should be sent into battle without at least an elementary degree of psychological self-understanding. He should never be permitted to infer that in some magic way he will not be afraid as he enters battle. Certainly he will experience fear, often marked fear, the natural protest of his strongest instinct and the most ancient biological function—self-preservation. Furthermore, he should not be taught that fear can be repressed or suppressed. It can no more be suppressed than can the beat of the heart be stilled. The soldier should be taught how to mobilize his resources so that he may learn to control his behavior when he is afraid. If this lesson is taught correctly and learned thoroughly, then fear becomes an effective fighting ally, motivating behavior that not only produces effective military action, but also gives the soldier the best chance of escaping with his life.

There are many other factors that go into the making of good soldierly morale, but if the few indicated could be thoroughly and constantly observed, unquestionably there would result a lowering of the incidence of the neuropsychiatric disabilities of war.

I have said nothing about the ideology and idealism of war. Should the soldier be given answers to this question: "Why are we fighting this war?" What part of the answers we might give will be inculcated into his personality and increase his morale? It will not be a very large part. Beyond the general conviction that our cause is more meritorious than that of the enemy, there is neither the time nor the place, nor does the average soldier have any desire or willingness to participate in forum discussions of political ideologies.

There are two beliefs that every soldier needs:

1. That the civilians at home are backing him, not only by buying war bonds, but in their attitudes and daily behavior. The morale barometer of the troops in the field dropped appreciably at the news of the coal and other strikes.

2. Every soldier needs confidence in the personality and quality of his leadership. It is true that an army good in its morale is apt to grumble but underneath, there is quiet confidence in their officers; belief in their capacity to look after them, to share common dangers with them; their participation with the officers in the knowledge of the objective to be taken; abiding faith that they will be led to victory.

"God, give us men. The time demands
Strong minds, great hearts, true faith and willing hands;
Men whom the lust of office does not kill;
Men whom the spoils of office cannot buy;
Men who possess opinions and a will;
Men who have honour; men who will not lie;
Men who can stand before a demagogue
And damn his treacherous flatteries without winking;
Tall men, sun-crowned, who live above the fog
In public duty and in private thinking."

TRENDS IN THE STUDY AND CONTROL OF INFECTIOUS DISEASES

RENÉ J. DUBOS

Professor of Comparative Pathology and Tropical Medicine, Harvard Schools of Medicine and Public Health

(Read April 20, 1944, in the Symposium on Wartime Advances in Medicine)

THE startling results which have been obtained by the use of drugs in the prophylaxis and therapy of certain parasitic and bacterial infections have been heralded by many as the dawn of a new era in the control of infectious diseases. Although we can indeed confidently expect that the near future will see great advances in the discovery and application of new chemotherapeutic agents, it must be emphasized that drug therapy constitutes only one facet of the complex problem of infection, and that its spectacular achievements and popular appeal should not lead to the neglect of the other aspects of the problem. The great strides which have been made toward the control of diseases caused by certain filterable viruses such as the agents of smallpox or yellow fever, bacteria like the typhoid and diphtheria bacilli, animal parasites like hookworms and schistosomes, were the result of painstaking analysis of the epidemiology of these infections, of enlightened public health practices, and of the skillful application of immunity reactions. Chemotherapy had no part in these important achievements. There is no reason to doubt that in the future, as in the past, the control of infectious diseases will have to be based on a clear understanding of their natural history and of the many factors which determine the outcome of the host-parasite relationship. In the following discussion an attempt will be made to recognize some of the channels along which the study of the infectious process is at present engaged or could profitably be directed.

THE NATURE OF VIRULENCE AND THE GENESIS OF EPIDEMICS

The discovery of the etiological agent of an infection is only one of the many steps in the understanding of the origin and in the description of the disease. The infectious process is the expression of a many-sided relationship between the pathogen and the host, and the complete analysis of this relationship requires a thorough study of its two components. It is for reasons of convenience only that we shall begin our discussion with a consideration of the parasitic agent.

Bacteria, and probably other micro-organisms, exhibit an extraordinary plasticity and can undergo profound variations of their biological properties. Thus, a given parasite can exist in different states of virulence and it is only when present in its most virulent form that it can give rise to disease and epidemics. Virulence, in fact, is an extremely complex property. In order to cause disease, the parasite must be able to reach the susceptible host; it must overcome the multiple defense barriers of humoral and cellular nature by means of which the body can rid itself of the many foreign substances that reach it; it must be able to multiply in the host and also to do damage to it. Each one of these properties is an independent attribute of the micro-organism which can vary independently of the others, and no parasite can establish an epidemic state unless it possesses all of these attributes at the same time. It will be sufficient to illustrate the complexity of the problem by considering some of the attributes which are required to render virulent the hemolytic streptococci of group A.

All streptococci isolated from pathological material are characterized by the possession of certain cellular constituents which can be demonstrated *in vitro* by the use of immunological reactions. Loss of these cellular constituents results in complete loss of virulence both for man and for animals, but, on the other hand, their presence is not sufficient to endow streptococci with maximal pathogenicity. In order to be fully virulent, these organisms must possess other subtle properties which have not as yet been identified with any known *in vitro* reaction; furthermore, these unknown properties are apparently lost under all sorts of conditions, and particularly when the organisms are cultivated in artificial media. Although the nature of these elusive attributes of virulence is not known, it is possible to define and characterize some of them in terms of their manifestations *in vivo*. Thus, a certain strain of

streptococcus can be highly invasive and cause a generalized bacteremia; it may produce a powerful erythrogenic toxin and cause intense scarlatina; one or more of its cellular constituents and products may produce severe febrile reaction and collapse of the host. Such highly pathogenic strains, however, may have only a low degree of communicability, or conversely, highly communicable organisms may produce little evidence of disease. All these factors vary independently of each other. To be capable of causing a severe epidemic, a given strain must possess several or all of the properties which have just been considered, and probably others as yet unidentified.

The analysis of the factors of virulence is not only a problem of academic interest. It contributes to a better understanding of the cause and nature of the disease and, consequently, to its more intelligent management. From the more special point of view with which we are now concerned, moreover, an understanding of the factors of virulence provides some essential information for the description and prediction of epidemics and of their course. Thus, although strains of streptococci or of meningococci are constantly present in men grouped in military establishments, the presence of those pathogens is not necessarily sufficient to give rise to epidemics. The latter result from transformations in the micro-organisms which enable them to spread rapidly from one individual to the other, from one station to another. What has been said of bacteria undoubtedly applies to filterable viruses, and, for instance, much remains to be learned of the circumstances under which influenza becomes epidemic.

There are, of course, many other factors which condition the establishment of an epidemic state. Among these can be mentioned variations in the general resistance and in the immunity state of the population, changes in the numbers and distribution of the vectors of the parasite, etc. Each one of these factors deserves the closest scrutiny and it is only for lack of time that they will not be considered here.

As more knowledge becomes available of the many factors which condition the epidemic pattern of an infection, it will become possible to forecast the course of epidemics instead of only describing them in statistical terms. We can foresee, I believe, the organization by public health authorities of "listening-posts" whose duty it will be to detect any qualitative and quantitative changes not only in the numbers of infectious agents present in the population, but also in those properties of these agents which are concerned with virulence. This indeed is not an idle dream. There are already functioning organizations, national and even international in scope, which follow the occurrence of diseases like plague, cholera, influenza, etc., in different parts of the world. Unfortunately, enough accurate knowledge is not yet available to recognize in most pathogenic agents these changes which are significant for the development of epidemics. In order to illustrate the nature of the problem, however, a specific example will be presented which, although it does not bear a direct relationship to the property of virulence, will serve to underline the importance of variability in micro-organisms.

It is known that bacteria susceptible to the sulfonamides can become resistant to these drugs when cultivated in media containing them. In fact, the development of drug-fastness has been recognized *in vivo* not only in experimental animals but also in human beings under treatment. There exists the possibility, therefore, that as a result of the widespread use of sulfonamides in therapy and especially for prophylaxis, there may develop in the population strains of pathogenic agents which have become resistant to these drugs. Although we have presented only a hypothetical possibility and although there is as yet no evidence of any real danger, the problem should not be ignored, and it is to be hoped that laboratories throughout the land will find it possible to maintain a permanent survey in order to follow the shift in susceptibility of the different pathogenic agents to the drugs in common use.

PROBLEMS OF IMMUNIZATION

The understanding of the natural history of an infection and of its epidemic pattern always suggests measures for its prevention and control. This has been the story of many of the enteric disorders and it explains why, on the whole, these infections have been so effectively dealt with in well-policed countries. It is certain, however, that preventive and public health measures are not equally applicable to all types of infectious diseases and that all these measures break down under the stress of emergency situations as, for instance, in times of war and other disasters. Whether preventive chemotherapy will ever become an advisable and effective practice is still a matter of conjecture. There is no doubt, on the other hand, that immunization can in many

cases become an effective supplement to public health measures, as has been proved in the case of typhoid fever. It appears worthwhile, therefore, to consider briefly the trends in the development of immunization procedures.

Ever since the introduction of smallpox vaccination by Jenner and the use of attenuated bacterial cultures by Pasteur, attempts have been made to produce a state of immunity in animals and in man by the injection of living attenuated avirulent germs. Vaccination against smallpox, yellow fever, plague, brucellosis are outstanding practical achievements of this method. There are, however, obvious objections to the use of a living vaccine since the danger exists that attenuated cultures can regain their virulence under unpredictable conditions. Most bacteriologists, therefore, have preferred to use vaccines killed by heat or antiseptics, as is done in the case of antityphoid vaccination. Furthermore, the use of killed organisms permits preparation of the vaccines from cultures of the highest possible virulence, and there is a widespread belief that the more virulent the culture used in the preparation of the vaccine, the more effective it is as an immunizing agent.

The theory that cultures of the highest possible virulence are essential to the production of high levels of immunity is based on considerations—not all of them entirely valid—which cannot be presented at this time. In any case, it is a fact that, in experimental animals, immunization with killed cultures of virulent pathogens can, under the proper conditions, elicit a high degree of immunity against the type of organism from which the vaccine was prepared. It must be emphasized, however, that this immunity exhibits an extraordinary specificity, since it does not protect against related organisms of another immunological type. Thus, mice immunized with a vaccine prepared from a certain type of pneumococcus, of streptococcus, or of dysentery bacillus are resistant to infection with bacteria of the same type but are still susceptible to infection with pneumococci, streptococci, or dysentery bacilli of another type.

Type specific immunity has lent itself to the preparation of effective therapeutic sera. For the latter to be used successfully, however, it is necessary that the specific type of the organism responsible for the disease be established by adequate tests, in order to permit the selection of the proper serum. The preventive immunization of whole populations presents an entirely different problem. Under practical field conditions pneumococci, streptococci, dysentery bacilli, etc., exist in a large number of types, many if not most of which are capable of causing disease; there is, furthermore, no way of predicting the relative prevalence of the different types in any outbreak. Since it is a practical impossibility to establish an effective level of immunity against all the different pathogenic types, there is reason to believe that immunization of threatened populations with type specific vaccines is an impossible goal. It is not unlikely that the practical success of antityphoid vaccination may be due in part to the fact that, in spite of minor differences, the strains of typhoid bacilli isolated from different outbreaks all possess essentially the same immunological constitution.

There exists, fortunately, a kind of immunity which transcends the limit of type specificity and which is effective against all the strains of one given bacterial species. Thus, it is possible to immunize experimental animals against pneumococcus infections under such conditions that the immune response is directed against a component of the bacterial cell which is common to all pneumococcus types. Similarly, one could probably find in all bacterial groups analogous cellular components, different from the type specific antigens and capable of giving rise to non-type specific protective immunity. Indeed, the effective immunity which follows the injection of attenuated non-virulent filterable viruses or bacteria may well be due to the non-specific protective antigens which we are discussing and which persists in the micro-organism even after the factors essential to virulence have been lost. It must be acknowledged that the level of immunity achieved by non-specific immunization is usually lower than that which results from the injection of type specific vaccines. Since, on the other hand, the attention of investigators has been focussed almost exclusively on the latter substances, it remains possible that intensive investigation directed in the proper channels would lead to the development of improved methods of non-specific group immunization.

Whatever the organism selected as immunizing agent, adequate techniques have to be devised to prepare from it an effective and safe vaccine. The methods which have been used heretofore are extremely primitive in their principle and consist essentially in killing the pathogen with

heat or antiseptics. The fact that in so many cases an effective state of immunity can be established by the use of killed bacteria or even filterable viruses indicates that the immune process is not directed against some mysterious living property of the bacterium or virus, but rather against some chemical constituent of its structure. In fact, it has been possible in a few cases to identify the particular bacterial constituents concerned in the reaction, and it is important to emphasize in this respect that of the many various substances which constitute a microbial cell only one, or a very few, are capable of eliciting the production of protective antibodies. Thus, it is certain that a very large percentage of the total material injected in antityphoid vaccination has no value whatsoever in establishing the immune state and contributes only unfavorable reactions. The separation from the parasite of these singular cellular components which give rise to protective immunity presents, therefore, great theoretical and practical interest. It is very likely that the use of purified antigenic preparations would tend to limit the untoward reactions which result from the injection of vaccines and would also permit a more accurate standardization of immunization procedures.

The preparation of purified antigens of bacterial origin suggests theoretical possibilities which haunt the dreams of the immunochemist. Since the immune process is directed against a well-defined substance of the bacterial cell and therefore against a well-defined chemical group, it is not beyond hope that the immunizing preparations of the future will consist of artificial materials designed to reproduce the immunological specificity of the pathogens which it is desired to control. Many steps must, of course, be taken before this aim is achieved; it is first necessary to recognize in the parasitic cell that particular substance which is concerned with the immunity reaction; this substance must then be isolated in pure form and its active radical identified before any attempt can be made to reproduce it by synthetic means. There is no prospect of any practical result in this direction in the immediate future, but the formulation of this goal serves to illustrate the view that, in the final analysis, the immune reaction is not directed against the infectious agent as a whole but rather against some of the specific chemical groups essential to pathogenic behavior.

THE DISCOVERY OF NEW CHEMOTHERAPEUTIC AGENTS

It is interesting that three of the most important steps in the growth of chemotherapy are the results of accidental unrelated observations, a fact which makes it difficult to read in the history of this science any logical predictable trend for its future. Quinine was introduced as a specific for malaria long before anything was known of the parasitic etiology of the disease; furthermore, the mechanism of action of the drug is as obscure today as it was when cinchona bark was first introduced in therapy. The sulfonamides were originally used in the treatment of disease in the form of a red dye, protonsil, which has since been found to be inactive against bacteria. Fortunately, protonsil is broken down in the animal body to sulfanilimide, the substance which we now know to be responsible for the chemotherapeutic activity of the dye. Much progress has been made during the past decade toward an understanding of the mode of action of sulfonamides, but the next great achievement of chemotherapy, namely, the discovery of penicillin, came from an entirely unrelated accidental observation. A mold growing as a contaminant in the laboratory was found to inhibit the growth of certain microbial species and to release into the medium a soluble substance, penicillin, which has the spectacular chemotherapeutic virtues which are familiar to all of you; nothing is known to date of the mode of action of penicillin.

There is one basic fact which must be kept in mind in any attempt to formulate a rational approach to the problem of chemotherapy. Although we know of so few drugs which retain their antibacterial activity *in vivo*, the bacteriologist has at his disposal a great variety of substances which exert a powerful antibacterial effect *in vitro*. These antiseptics can be produced by the methods of synthetic chemistry or by a number of microbial agents, molds, actinomycetes, or bacteria. It is clear, therefore, that the difficulty is not in finding more and more antimicrobial substances, but that the real problem is rather to define that property or combination of properties which permit an antiseptic to retain its activity in the presence of animal tissues, to behave, in other words, as a chemotherapeutic agent. Although the facts known at the present time are too few to permit a statement of the problem in general terms, it is certain that the most effective chemotherapeutic agents do not behave as gross protoplasmic poisons which

destroy the metabolic activity of all living cells, but rather as selective inhibitors of some specific steps concerned in the nutrition, synthesis, or cell division of the parasites.

It is beyond the scope of the present discussion to consider the avenues of approach which are most likely to lead to the discovery of drugs useful in the treatment of infection. One may suggest, however, that the search for new chemotherapeutic agents should not be limited to the mere production by the methods of organic chemistry or from natural sources of more and more bacteriostatic and bactericidal substances, but that efforts should be made to determine what properties of a substance permit it to retain its antibacterial activity *in vivo* without causing irreversible damage to the host. Progress in this direction will depend in part upon an increased knowledge of the steps of microbial metabolism for which there can be developed specific inhibitors exhibiting selective affinity for the parasite.

THE RESPONSE OF THE HOST TO THE PARASITE

There is no apparent reason why the mere presence in the tissues or body fluids of any reasonable number of parasites of microscopic dimensions should exert any harmful effect on the host that harbors them. No mechanical theory of the pathogenic action of bacteria or filterable viruses is compatible with our knowledge of the way in which the tissues deal with the particles which have gained access to them. In the final analysis, the harmful effects of most infectious agents are chemical in origin. It is through a disturbance of the normal physiological processes of the host that pathogenic organisms cause those symptomatic and pathological manifestations which characterize each individual infectious disease. In some cases it is possible that the parasite interferes with the normal physiology of the host by competing with the latter for some factor essential to vital processes or by producing simple metabolic products which alter the course of the biochemical events of normal metabolism. These possibilities have been so little explored that they can be stated only in the most general terms. In most cases, however, it has been established that pathogenic bacteria produce a variety of substances—the bacterial toxins—which are endowed with great pharmacological activity and which are undoubtedly responsible for many of the clinical and pathological manifestations of each infectious agent. It would be of the greatest interest, consequently, to understand how toxins exert their physiological and pathological disturbances, since only then shall we gain an insight into the intimate mechanism of infectious diseases. Unfortunately our ignorance of these problems is truly appalling. Toxin action has been analyzed and described only in terms of the secondary phenomena resulting in pathological and clinical manifestations. Much is known, for example, of the symptomatology and pathology of the toxaemia caused by the diphtheria, tetanus, and botulinus toxins, but we know nothing of the primary biochemical or physiological lesions for which these toxins are responsible, whether they act by destroying tissues or cellular structures, or by inhibiting essential metabolic functions. Knowledge of the initial reaction which takes place between the toxin and its susceptible substrate in the body would permit a more accurate definition of the specific physiological functions which are altered during the infectious process. This knowledge would contribute to the understanding of disease and would in the long run lead to the development of specific therapeutic measures effective against toxaemia.

Our complete ignorance of the physiological basis of infectious diseases is due in part to the fact that, since the beginning of the microbiological era, the study of infection has been almost exclusively limited to the immunological aspects of the host-parasite relationship. It is obvious, however, that not only the severity and the character of the disease, but even its very occurrence, depend as much upon the reaction of the host as it does upon the presence of the infectious agent. Familial susceptibility to tuberculosis, to rheumatic fever, etc., is a well-recognized phenomenon, which has been confirmed experimentally by the breeding of selected lines of animals either highly susceptible or highly resistant to a given infectious agent.

It is also known that obvious nutritional deficiencies can modify the susceptibility to infection and there is reason to fear, for example, that an increase in the morbidity and mortality of tuberculosis will follow in the wake of the war in Europe. On the whole, however, very little is known of the relation of nutrition to resistance. Nor do we have any significant information concerning the effect of hormonal regulation on the course and outcome of the infectious process.

The nature of the anatomical and physiological characters of the host which condition its response to the parasite is one of the most neglected and most important aspects of the problem of infection.

Because of the lack of the most elementary information, it is not possible to predict whether an understanding of the physiological factors of host resistance could ever be translated in terms of practical therapy. One can hardly doubt, however, that this knowledge would help and guide the physician in the understanding and treatment of the patient. Thus, it is the present-day practice based upon a general statistical experience to advise any individual showing evience of incipient minimal tuberculosis to follow a severe routine of absolute bed rest. It is well known, on the other hand, that the course and outcome of tuberculosis varies greatly from one individual to another and there is no doubt that the management of the individual patient would be greatly influenced if it were possible to predict the course of his disease under a given set of conditions. In other words, the whole policy of sanatorium treatment could be restated in more rational terms if it were possible to predict by adequate tests the expected response of a given individual to the tuberculous infection.

The course of any science is often conditioned more by the availability of easy experimental methods than by the relative importance of the potential lines of investigation. There were developed early in the microbiological era a number of techniques based on serological and immunochemical reactions which are so rapid in their performance and which have indeed yielded such important and useful information that they have directed the study of infectious diseases along somewhat narrow channels. It is likely that, by returning to the main channels of biological and biochemical philosophy, the student of infection will achieve a more complete picture of the many reactions by which the host responds to the specific stimuli exerted by the parasite, and will thereby devise novel methods for the control of infectious diseases.

R. A. F. PENROSE, JR., MEMORIAL LECTURE

THE INTERNATIONAL LABOR ORGANIZATION LESSONS FOR TWENTY-FIVE YEARS

HAROLD B. BUTLER

Minister from Great Britain to the United States

(*Read April 21, 1944*)

It is a rather curious feature of present-day discussion of the post-war world that so little attention is paid to the experience gained in the last twenty-five years. The League of Nations and all its works are usually brushed aside. It is held to have failed, but little effort has been devoted to enquiring why it failed and how the causes of its failure can be avoided in the future. This tendency to ignore the League experiment in international organization would be more easily understood if the lesson drawn from it was that all international organization was impossible or undesirable. Yet this is not the conclusion towards which opinion in the United Nations seems to be moving. On the contrary there seems to be a growing conviction that some sort of international organization is necessary, indeed inevitable, if after the present cataclysm peace is to be restored on a stable basis and the world reconstructed on progressive and prosperous lines. As early as 1941 the Atlantic Charter foreshadowed some measure of collective action to secure the maintenance of peace and collaboration between nations in the economic field. The two major attempts which have already been made to tackle the most immediate problems of reconstruction have resulted in the erection of two important pieces of international machinery —the Food and Agriculture Council and the United Nations Relief and Rehabilitation Agency. Only a few months ago the Moscow Conference issued a declaration which recognized the necessity of setting up "a general international organization . . . for the maintenance of international peace and security."

From all this it may be safely inferred that the League of Nations and its sister institutions are not regarded as unqualified failures. Their experience appears to have carried the conviction that, whatever their defects and shortcomings may have been, the solution of the world's major problems must be sought not in a reversion to an anarchical state of international society, in which every nation was against its neighbor, but in an effort to integrate them into a community by means of some sort of collective institution for the purpose of holding it together and of devising common policies for its general guidance. I therefore make no apology for examining briefly the history of one of the League institutions and for trying to suggest some of the lessons which may be derived from it. Having attended the birth of the International Labor Organization at Paris in 1919 and having lived and worked with it continuously for nineteen years afterwards, it naturally gives me a peculiar parental pleasure to see that it has weathered the great storm and is now vigorously preparing to play its part in the dark and difficult days which lie before us. It is a good omen that it should be setting its new course from Philadelphia, whose name and whose history pre-eminently befits it to bless this new venture in human fraternity.

Perhaps the most surprising thing about the I.L.O. is the fact that its constitution has withstood all the strains and the stresses of its first twenty-five years so well as to need little amendment. Its foundations were more truly laid than we could have dared to suppose in those feverish, uncertain days of the Peace Conference. Time has shown, and has shown with increasing emphasis, that its basic conception responded to a vital need of our times. It was founded on the belief that the improvement of the lot of the common man must be one of the main objectives of human society, that there could be no peace in the world either within nations or between nations, unless the grosser inequalities and injustices which have marred human relations were progressively eliminated. That belief was only dimly dawning twenty-five years ago. It had not been generally accepted as an axiom by which national policies must be judged. During the intervening time, however, its truth and necessity have been more and more widely recognized, so that today what seemed a Utopian

vision has now been accepted almost as a truism by all those who believe in democracy. In other words the course of events has gone to prove the rightness of the assumptions upon which the I.L.O. was founded. The successive crises through which the world has passed—the chaotic recovery of the early twenties, the great slump of 1929 which dispelled the short-lived and specious dream of prosperity preceding it, the economic anarchy and the widespread unemployment which accompanied it in the early thirties, the social and political upheavals which they produced in every continent—all these brought into grim, vivid relief how much the fate of the ordinary man and woman was the plaything of vast, unseen forces, against which their feeble individual struggles were condemned to futility and frustration.

During its earlier years the efforts of the I.L.O. were mainly devoted to the construction of an international code of labor standards, and that essential work has gone steadily forward ever since. Norms were set by international treaty for hours of work, the protection of women and children, the prevention of accident and industrial disease, the development of social security to fortify the individual against the onset of unemployment, sickness and old age. In all these fields steady progress was recorded year by year, as ratifications registered in all the five continents flowed in to Geneva. A world-wide network of obligations was gradually woven. The result has been to make the lot of the worker in industry in most countries easier and less hazardous than it was a generation ago. And this improvement was not just a matter of chance or caprice; it was securely based upon formal engagements taken by nations to maintain it. Nor were these benefits confined to those who earned their livelihood in the industrialized countries of the west. As industry spread to the East, India and China began to model their practice upon the standards set by the I.L.O. As might be expected, their benefits were more keenly felt in the less advanced countries than in the older industrial countries with a long history of social legislation and trade-union activity. The agricultural worker also began to receive his share of attention. The conditions of labor in colonies and dependencies were brought under strict regulation by conventions abolishing forced labor, penal sanctions, and abuses in the recruitment of native labor. Not only in Europe and the Americas, but in Asia and Africa also there was hardly a territory or a community which had not received some ray of light and encouragement from the first attempt to tackle the social problem systematically by a great collective endeavor. That was what the work of the I.L.O. implied. Its name had already become a legend as I often found on my travels, not merely in the great centers of industry in America and Western Europe, but in the small towns of Egypt, the factories of India and Malaya, and even vaguely in obscure villages in the heart of Africa.

But all this, good and beneficent as it was, did not reach down to the roots of human well-being. Every system of social advancement rests on two tacit assumptions, first that no state of war exists, secondly, that economic balance is reasonably maintained. Of war it is unnecessary to speak. Under its categorical imperative all social standards are apt to go by the board. Hours of work are long, vacations are sacrificed, the ordinary safeguards against overwork and fatigue are jettisoned in the interest of national defence. These things are accepted as a matter of course, but the obvious conclusion is not always drawn—that some organized system for the insurance of world peace is the indispensable foundation of social progress. To go blithely forward with ambitious social planning without first taking care that it will not be swept away by a furious blast of destruction in another ten or twenty years' time is surely the height of folly. The motto of the I.L.O. is that there is no peace without social justice. The converse is equally true. There can be no social justice without peace.

But in order to give social progress a favorable prospect, the prevention of war is not sufficient. Economic upheaval is hardly less fatal to social progress than political upheaval. Without a stable economic foundation it is built on sand. Social security is of precarious value to those whose whole existence is threatened by the loss of their jobs and of their standing in society. Shorter hours of work are a mockery for those who have no work. Insurance against unemployment or sickness is only a palliative for those whose incomes, whose savings, and whose homes have been swept away by some economic blizzard, blowing no one knew whence, driving millions of people before it into sudden and irretrievable disaster. That was the hard lesson of the great slump of 1929–1932. Though it is now a matter of history, it is not forgotten history. It changed the whole outlook of mankind on the

social problem of our time and bred a determination which still persists to prevent the recurrence of such catastrophies. Eleven years ago I had the honor of writing an introduction to a monograph on the I.L.O. published in the *Annals of the American Academy of Political and Social Science*, which closed with these words:

> As the world becomes more closely knit, its economic destinies are more and more the common lot of all its citizens, without distinction of nationality. It is only through co-operative planning and action that disasters like the great depression can be avoided in the future. That is the most inescapable lesson of the crisis; and it is for that reason that the international economic machinery established at Geneva, of which the International Labor Organization forms an increasingly important part, must grow rather than diminish in importance if the recurrence of economic catastrophies is to be prevented.

I believe not only that those words are still true today, but that their truth has become widely recognized. Social progress cannot be dissociated from economic stability. The one cannot exist without the other. In a world subjected to violent fluctuations of production and employment there can be no security for the individual. Without individual security there can be no solid basis for social life. During the last fifteen years no great industrial country has escaped internal troubles directly traceable to the seething discontent which is the infallible product of widespread unemployment. That has become the problem which haunts the thinking of every government. No nation can allow millions of its citizens to tramp the streets in enforced idleness without running the risk of political convulsion. For the moment war has provided a drastic solution. By drafting millions into the armed forces and by pouring billions of dollars of public money into the production of raw materials and war materials, the belligerent countries have temporarily banished unemployment and created a condition of artificial prosperity, which affects every country outside the actual theatres of hostilities. But that condition cannot last. When the United Nations have finally crushed the gangster powers, when the enslaved countries have been liberated from their tyranny, peace will come again, and with it the problems of peace in a new setting. The world will be confronted by a terrible spectacle. Not only will millions have perished on the field of battle, in the bombed cities and under the machine-guns of execution squads, but the fields and factories of great portions of Europe and Asia will have been laid waste. Even in countries unscathed by war other millions will have temporarily lost their means of livelihood until the vast machine of industry has been reconverted to civil production. And when that conversion has been effected, how far will an impoverished world be able to buy what industry will then be geared to sell? Until the farmers and the manufacturers of Europe can produce, how will the great European market with its three hundred million people be able to pay for the goods which the other continents can offer it? Between 1929 and 1932 the world index of industrial production fell from 110 to 65, not because the great masses of people did not need all and more than industry and agriculture were able to produce, but because they had not the purchasing power with which to buy. That problem is going to recur, perhaps in an aggravated form, after the present struggle, which will have obliterated far more of the world's accumulated wealth and capital equipment, its houses, machines, plants, and man power than the First World War of a generation ago.

This tremendous situation cannot be met unless another lesson of the great slump is applied. If anyone had doubted the economic unity of the world before, the great depression must have disabused their minds. It respected no type of national economy and no type of political constitution. It smote agricultural countries and industrial countries, democracies and dictatorships with the same ruthless and impartial violence. Workers and industrialists, stockbrokers and stockholders, farmers and miners went down together before its blast. Its deadly effects were felt by the steel plants of Pittsburgh, Sheffield, and Essen, the textile mills of Lancashire, Bombay, and Massachusetts, by the cattle ranchers of the Argentine, the coffee-growers of Brazil, the sugar-planters of Java and Cuba, the rubber-planters of Sumatra and Malaya, the cotton-growers of Tennessee, India, and Egypt. None was spared. A creeping paralysis seeped into every corner of the economic world. And yet no universal effort was made to combat this universal disease. Each nation struggled desperately and alone to find its own salvation at the expense of its neighbors. Instead of a concerted attempt to ease the processes of buying and selling and to facilitate the exchange of goods and services within nations and between nations, energy under the spur of panic was directed to blocking the channels through which the tide of

renewed prosperity might flow in to revive the world's sick economy. Loans were restricted, money-rates were raised, tariffs, exchange controls, quotas, and subsidies were piled up in a desperate effort to promote recovery. The volume of world trade fell by 25 percent in three years and had not recovered its 1929 level when the great war broke out ten years later.

If then we are to realize the high hopes of a better world without which the outlook for the future would be bleak indeed, we have got to tackle the social and the economic problems together. They are simply the two faces of the same coin. They are just two aspects of the same process of reconstruction, in which, as the agenda of its Philadelphia Conference testifies, the I.L.O. must play a large part. It rightly claims "the right to examine the economic policies that are being pursued in order to ensure that they will in fact serve the purpose of promoting full employment and a rising standard of living throughout the world." Unless those vital aims are achieved, the efforts of the I.L.O. to improve the lot of the worker will be condemned to frustration. They cannot be achieved by one nation or even by all nations acting independently and discordantly. The economic unity of the world is the plainest lesson of the last twenty-five years. Unless it is taken to heart and applied by working out common policies for the restoration of the world's economic health, our last state will not be better than the first.

And the argument leads us a step further than this. Unless we can solve the economic problem, we shall in the end discover that we have not solved the problem of peace. Just as economic stability is impossible without political stability, so peace is impossible without a well-integrated economic system, which will enable the production and exchange of goods to go forward on an expanding scale all the world over. Economic warfare is apt to become the prelude to military warfare. Unless economic antagonisms can be eliminated, unless we can realize that the prosperity of our neighbors is essential to our own prosperity, we shall not achieve peace between nations any more than we shall achieve social justice within nations. The political, economic, and social problems are not three separate problems, but different aspects of one and the same problem—the supreme problem of preserving and developing a good civilization as we conceive it.

I do not need to labor this point, but, to clinch it, I should like to quote a few words from Professor Carl Becker's remarkable book, *How New Will the Better World Be?* which has very recently appeared. He says that whether an international settlement amounts to anything in his view, "depends upon what measures are taken to solve fundamental economic problems. Let the United States, Great Britain, Russia, China, and the minor states 'co-operate' with the best will in the world. Let them set up whatever political league, federation, or 'international machinery' they like, backed by whatever 'international police force' they think necessary. They will still find that to prevent war and aggression, to abolish 'fear and want' it will be necessary to do more than get rid of a bunch of 'war-breeding gangsters.' They will find that it is necessary to deal with such matters as international trade and investments, currency stabilization, the development of industrially backward countries, varying standards of living in the several countries, and access 'on equal terms' to the raw materials of the world needed by each country. But such problems cannot even be intelligently discussed, much less solved, without taking into account the measures adopted in each country for dealing with full production and unemployment, tariffs, foreign investments, currency regulation, the control of patents and the activities of corporations involved in international cartels, and the like. If there is to be a new and better world, the international economic order must be more or less planned; and national and international economic problems are so intimately related that planning in either field cannot very well succeed unless it is in some measure integrated with planning in the other."

But even wise planning and far-sighted policies are not enough, unless they are inspired by a true conception of life. Without that war will not cease nor will prosperity return. Neither of these goals can be reconciled with the fascist spirit. During the inter-war period that spirit was in constant conflict with the endeavor of the I.L.O. to promote a more humane and spacious way of life for the common man. The essence of the fascist philosophy is that the supreme objective is not the welfare of the individual but the power of the state and of its ruling clique. To that end every amenity of life must be sacrificed, every activity of the common citizen directed, regardless of the consequences to his happiness and well-being. It was summed up in the well-worn but accurate formula "guns

instead of butter," though it may be noted in passing that in all the fascist countries the ruling clique saw to it that they had plenty of butter as well as the power to use the guns. It is sometimes supposed that this antithesis is the same as the antithesis between dictatorship and democracy, but this is not necessarily the case. The true criterion is not one of constitutional structure, but of ultimate objective or what the Germans call "Weltanschauung." There are infinite degrees and gradations of democracy just as there are monarchies and dictatorships of many shades and colors inspired by very different objectives and responsive in very different degrees to the desires of the people. There are many countries whose political and educational development do not fit them for the more advanced forms of democratic government as practised in North America and Western Europe. It would be foolish to suppose that European countries new to democracy, the semi-barbaric tribes of Africa or the untutored masses of Asia were capable of exercising political responsibility in the same measure as the citizens of the United States, Canada, Britain or Holland or France or Australia with the habit of forming political judgments for themselves created by generations of political experience. There are many stages on the road to full representative government. The real test, however, is to be sought not so much in the form of its political institutions as in the aims which the state pursues, and the degree to which public opinion endorses them. The criterion which has to be applied is whether its objectives are those pursued by the military oligarchies of Germany and Japan, or whether its efforts are primarily devoted to raising the standard of life of its citizens by the spread of education and culture, by the development of its natural resources, by increasing the yield of its agriculture and by developing industry not for the purpose of producing armaments, but in order to supply the needs of its people.

During the first years of its existence, the I.L.O. was carried forward on a wave of democratic sentiment, on the belief in the freedom and happiness of the individual as the supreme ends of political and social activity. That concept had its roots in the English, American, and French Revolutions. Throughout the Nineteenth Century it had been enthusiastically or reluctantly accepted even by monarchs and dictators, however much their practice may have at times belied their professions. But since the last war the very idea of popular progress has been bitterly and openly challenged first by Italy, then by Japan, and finally and most ferociously by Germany. They claimed that both government by the people and government for the people were utterly false philosophies. In opposition they raised the ideal of the power-state exacting discipline, efficiency and military preparedness for their own sake, regardless of their effects on the lives of individuals. Instead of deriving their right to govern from the will of their citizens, they only conceded to them such rights and imposed on them such duties as might seem good to the rulers. They denied any right to education or culture of self-determination which did not further the march of the state to power, power over its subjects within, power against its neighbors without. That is what we are fighting against. That is why all hopes of social progress are vain, unless we first win this war.

But when this war is won, the fascist danger will not have been finally exterminated. Its poisonous seeds have been sown far and wide. To root them out will require eternal vigilance and a firm determination to prevent them from ever coming again to flower. That, as I see it, is one of the tasks of the I.L.O. From 1933 onwards, one fascist state after another repudiated it in open defiance or tried to undermine it from within. Its very existence depends upon the postulate of individual progress as the supreme end of human society. Only insofar as we remain faithful to that ideal handed down by our forefathers can we look forward to a period of peace, prosperity, and progress. Each is unattainable without the other two. To reach them we have no hope as individuals or as nations unless we work together with a common purpose in a common enterprise. That is the final and inescapable lesson to be drawn from the failure of the first attempt at international organization during the past twenty-five years. The future will show whether we have the vision to comprehend it and the ability to apply it.

We are faced by a formidable task. When the war ends we shall find ourselves in a world more distracted, devastated, and dislocated than it has been for centuries. It will be a very sick world. To suppose anything else, to imagine an easy return to a normal comfortable existence is to harbor a most dangerous delusion, which would blind us to the true nature of our situation. The process of recovery will at best be very slow and

infinitely painful. Clear thinking and controlled action will be very difficult amid the welter of misery, frustration, and confusion in which millions of human beings will be floundering. To evolve order out of this chaos may well seem a superhuman task, but it will only become so if we lose faith in ourselves. If we hold fast to our belief in the best in human nature, we can go forward bravely to the conquest of the huge obstacles in the path of reconstruction. If, however, we despair of a better world, if through weakness or weariness of spirit, we resign ourselves to abandoning the hope of it, then we shall be condemned to failure which may project our children into an age of deepening darkness. As that great man General Smuts said, "there is no malign fatalism which makes fools of us in our dark striving toward the good. On the contrary, what is highest in us is deepest in the nature of things." That is the only precept which offers salvation. If we follow purely egoistic and materialistic aims, we are sure to be disappointed. If we ignore the good of our neighbors, we shall cheat ourselves of our own. If we put economic above spiritual values, we shall fall into inevitable decline. If we give way to the cynicism of despair, we shall reap as we have sown. But if we keep our ideals bright, if we cling steadfastly to the belief that humanity by its own exertions can attain a higher and nobler destiny than anything it has known, then we shall not fail. Hopes may be dupes, but fears also may be liars. Many of our hopes of the last twenty years have ended in disillusionment, but our fears for the future will be dispelled if we meet them boldly with unfaltering faith in the principles of freedom and with the courage that is always given to the true worshippers of liberty.

MOUNTAIN GROWTH, A STUDY OF THE SOUTHWESTERN PACIFIC REGION

WILLIAM HERBERT HOBBS

Professor Emeritus of Geology, University of Michigan

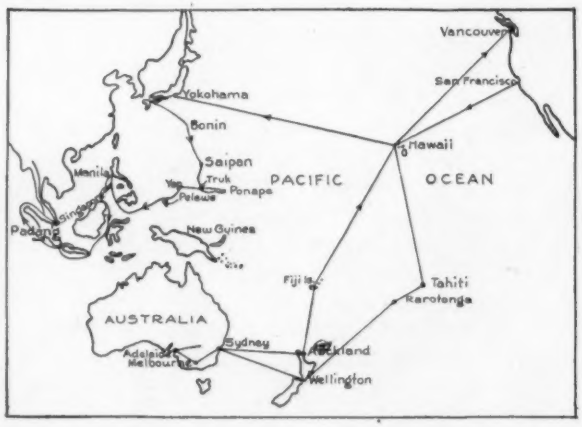
(*Read April 23, 1943*)

ABSTRACT

In most parts of the world where mountains are to be found, the history of their origin and manner of growth is difficult to decipher, and mainly for the reason that the history, if not already completed, is at least in a very advanced stage. We are in those areas dealing with conditions of erection which no longer obtain, and throughout this history other and destructive rather than constructive agencies have been in operation, and these agencies have largely effaced or else entirely removed the pertinent evidence. These agencies are grouped under the names weathering, erosion, and transportation, and they have left us only disfigured remnants of forms and structures which at one time would have laid bare the essential facts of the history.

These destructive agencies are largely within the realm of the atmosphere, and it is within that shell of our planet that most mountains undergo their erection.

In one region, that of the southwest Pacific, and in that alone, mountains have in very recent time been rising from the bottom of the deep sea, where they have been immune from the usual destructive agents, and they are there even today in an early stage of their growth. Impressed by the unique opportunity for study which was there presented, the author in 1918 began a preliminary study of the area, and in 1921 and 1923 he undertook for the purpose extensive reconnaissance cruises in the western and southwestern Pacific areas (map 1).



MAP 1. The author's cruises in the Pacific region in 1921 and 1923.

The proof that the mountains within this southwestern Pacific region are still in a vigorous growth which is not duplicated elsewhere, is supplied by the authoritative map of earthquake distribution (map 5). Only less important as evidence is the local effusion of now active volcanic vents which reveals general elevation while the presence of coast terraces reveals a local elevation, as barrier reefs do local subsidence. A very important aid has also been recently supplied by the extended use of the echo method for deep sea soundings (map 7).

The present study is treated in three parts: I, a general discussion of the nature of mountain growth; II, a special study of the mountain arcs of the southwestern Pacific area; and III, a study of the principal islands in the oceanic area of subsidence where crustal settlement has brought about the wrinkles or arcs.

CONTENTS

PART I, GENERAL

1. The ultimate and the proximate causes of mountain growth 222
   The ultimate cause—a shrinking of the earth's core 222
   The ocean floors the areas of subsidence 223
   The evidence for the depression of the Pacific floor 223
   Marginal mountains a consequence of the settlement 224
2. The underthrust outward from oceanic settlement . 225
   Underthrust landward as against overthrust seaward 225
   Compression of a mountain arc during its erection 226
   Resolution of the underthrust at the bottom of the continental slope 227
3. The pattern of a system of mountain arcs 227
   The relatively rigid coigns within the central areas of the continents 227
   The latest arc system borders the Pacific Ocean . . 228
4. Mountain arcs are flexures with a syncline on the seaward side of the anticline 230
5. Active volcanoes accompany the erection of arcs . . 233
   Volcanic vents on the back of the arc 233
   Initial lava effusions have the composition of mudstone—andesite 233
6. Theory of origin of the volcanic effusions 235
   Fusion of rock beneath a competent arch 235
   Lava rises to the surface through the continued compression of the anticline 236
7. Evolution in the plan of the arc 236
   Earlier stages of relatively flat curvature with trough at the front 236
   Rising arcs the locus of strong earthquakes 239

PART II, MOUNTAIN ARCS OF THE SOUTHWEST PACIFIC REGION

1. The Bonin arc 239
2. The Iwo arc 240
3. The Marianne arc 241
4. The Yap arc 244
5. The Palau arc 246

6. The Sonsol arc 250
7. The Riukiu (Loochoo) arc 250
8. The Philippine arcs 251
9. The Sunda-Banda arc system 254
10. The arc succession in the Philippine-Sunda-Banda arc system 255
11. The Bismarck Archipelago—Solomons—New Hebrides arcs 256
12. The Tonga-Kermadec arcs 258
13. The New Zealand arcs 259

PART III, THE SINKING PACIFIC FLOOR

1. The lava effusions 260
2. The partition line between mountain arcs and the sinking floor 261
3. Basaltic vents in the southwestern Pacific floor .. 262
4. Truk 263
5. Ponape 264
6. Kusaie 266
7. Summary and Conclusions 268

NOTE: Until World War II the islands of the Pacific had been vouchsafed comparatively little attention in American circles, as their importance commercially and strategically had been but lightly regarded. For purposes of description they now offer a difficult problem because of the many different names which have become attached to the same island and the different spellings of each which are in use. The multiplicity of names came about partly because of duplication in discovery, most frequently by unlettered whaling or other sea captains. Many of the names are of native origin and they have been phonetically rendered in several different languages. Though the discoverers have more often been either English or American, the superior atlases published on the continent of Europe and in general use for reference in our libraries, have brought much German or French spelling into our practice.

One or two examples may be mentioned. The Mariannes were discovered by Magellan and named by him the Ladrones (Thieves) because of his experience with the natives. The islands later took the Spanish name *Marianas*, but into wide use has come the word *Mariannes* (either French or English) and this form I have used. The Pelews or Pelew Islands were given the name Arricifos in 1543 by their discoverer, Villalobos. They were later given the English name Pelews, and still later sometimes Palau, perhaps because so many German scientists wrote monographs on them.

The first attempt to bring some order out of chaos was made by William T. Brigham, Director of the Bishop Museum at Honolulu, who in 1900 brought out his "Index to the Islands of the Pacific" (*Mem. B. P. Bishop Mus.* 1, no. 2; 170 quarto pp., maps). His usage I have rather generally followed.

World War II at once focussed the world's attention upon the Pacific region and, all too late when my maps and figures had been drawn, the National Geographic Society issued its superior series of Pacific maps. These maps, and the indexes to them which are now issuing, have been made standard by the approval of our army, navy, and air forces. Rather generally they adopt the original spelling, and so our Porto Rico becomes Puerto Rico.

PART I, GENERAL

1. THE ULTIMATE AND THE PROXIMATE CAUSES OF MOUNTAIN GROWTH

THE ULTIMATE CAUSE—A SHRINKING OF THE EARTH'S CORE

The lithosphere or outer rock shell of our planet seems to betray by its oblate spheroidal form, when this fact is considered in connection with its still continuing axial spin, an origin in a plastic condition. It is generally held that this early plastic spheroid through continued loss of heat into the surrounding space has become solid at least in the near-surface shells.

That the earth is now rigid throughout its mass, and probably also in a non-fused condition, would appear to be indicated by at least three observed conditions. These are:

1. The lithosphere is without generally appreciable tides like those of the sea. As measured by Michelson and Gale, the earth tides due to the pull of the sun and moon are of such small amplitude that they can be revealed only with use of the interferometer.

2. If the earth had a molten interior, it must, as reasoned by Kelvin, have long since ceased to spin on its axis.

3. The velocity of compressional earthquake waves which traverse the earth's interior has been repeatedly measured and found to be comparable to that of such waves traveling in the best tool steel. Confirmation of this is found in the period of precession or tilting of its plane of spin—precession of the equinoxes.

But if the earth's interior is not in a state of fusion, it does not follow that interior temperatures are not sufficient to melt rocks if they were under surface conditions of pressure. The outer rock shells are under high downward pressure (load), and it has been established that the melting point of all those rock types which have been investigated increases at a rate greater than that of the pressure from load.

The issue of lava from volcanic vents proves only that pockets (*maculae*) of molten rock exist at some moderate depth below the surface. If the geothermic gradient, or increase of earth temperature with depth that is found to obtain for the first few miles below the surface, should continue to as much as fifteen miles, all rocks would under surface conditions be fused even if in a dry state, and because of the water present at considerably lesser depths.

Wherever rocks below the earth's surface are at fusion temperatures, but maintained unfused and rigid by the load upon them, if this pressure of load should be locally removed, they would obviously become fused and yield a pocket or *macula* of *magma* (lava), and so supply a volcanic vent at the earth's surface.

Since, then, the earth's subsurface shells are doubtless at or above the surface fusion temperatures of rock (800° to 1500° C.), and since they lose heat into surrounding space, the earth must continue to shrink. Because of this contraction of volume, portions of the surface will move inward toward the center of form, and this will result in a reduction of the superficial area of the planet. The areas which have withdrawn most toward the earth's center are obviously the ocean bottoms, and from these great areas of settlement thrusts will be exerted outward toward and into the coastal areas of the continents.

The continually augmenting stress condition within the coastal regions of the continents is at irregularly recurring intervals relieved in great spasmodic mass jolts, earthquakes, and the face of the country is thus materially changed. At such times the behavior of all structures which are continuous, such as rails, pipes, curbs, wires, bridges, etc., are looped or buckled in such a way as to prove that a reduction in superficial area has occurred. There are no exceptions.[1]

THE OCEAN FLOORS THE AREAS OF SUBSIDENCE

Proof that the ocean floors were once farther removed from the earth's center of form, and have descended to their present positions as a result of a settlement, or subsidence which is measured in many thousands of feet, seems to have been afforded by the formation of encircling or barrier reefs about oceanic islands and by the atolls into which many hundreds of them have evolved.

THE EVIDENCE FOR THE DEPRESSION OF THE PACIFIC FLOOR

Quite independently the evidence for the depression of the oceanic floors during late geological time was pointed out by Charles Darwin and by James D. Dana, as a result of studies made by each when on long world cruises; by Darwin on H. M. S. *Beagle*, and by Dana on the United States (Wilkes) Exploring Expedition. The evidence as given by Darwin was illustrated by him in the now classic successive diagrams of figure 1.

FIG. 1. Diagrams to illustrate the formation of encircling (barrier) reefs and atolls as a result of a subsiding ocean floor. (After Darwin.)

In all warm seas where the temperature of the water does not fall below 68° F., free-swimming, reef-building organisms, both animals and plants, attach themselves to the rocks of island shores and form fringing reefs (diagram at the left in fig. 1). All oceanic islands far from the continental shores are volcanoes, and most of these have or have had fringing reefs. The growth of the reef-builders is inhibited by water pressure at a depth much greater than that at 20 fathoms (120 feet).

If now a settlement (or a series of settlements) of the island takes place which does not exceed 120 feet, the reef colonies will continue to grow, and in the course of time the reefs will arrive at the surface as an encircling reef separated from the shore of the island by a ring-shaped lagoon (middle diagram of fig. 1). By repetition of such subsidences the volcanic island later becomes submerged, and the encircling reef is changed into an atoll—a rim of reef enclosing only a lagoon (diagram to the right in fig. 1).

Atolls are found by the hundreds distributed widely throughout the warm seas, and if the fundamental assumptions of the Darwin-Dana theory are warranted, there has been a settlement of the Pacific floor since late Cretaceous time which amounts in the aggregate to at least 10,000 feet. Since the theory was violently attacked in favor of a rival explanation of formation of atolls, offered by Sir John Murray and defended by a group of zoologists and notably by Alexander Agassiz and J. Stanley Gardiner, it is best to consider here briefly the results of a test to which the rival theories were subjected by the Royal Society of London. It had been early pointed out that since, according to the Darwin-Dana theory, the atoll evolved about a volcanic mountain of quite definite form, a boring put down through the atoll, if carried far enough, should

[1] Hobbs, W. H. A study of the damage to bridges during earthquakes. *Jour. Geology* **16**: 636–653, 1908.

——Studio dei danni prodotti dai terremoti ai ponti (Giulio Fornari, transl.). *Ingegneria Ferroviaria* **9**: 1–20, 1909.

——Construction in earthquake countries. *Eng. Mag.* **37**: 929–947, 1909.

enter the volcano at its base. A committee of the Society was set up under the chairmanship of Sir John Murray, funds were provided, the atoll of Funafuti was selected, and an expedition was sent out in a naval vessel to put down the borehole and obtain a core. At the time (1896) the oil industry had not brought the technique of rock borings to the perfection it has since attained, and after much difficulty the bore was carried to a depth of only 1,114 feet. Since it did not penetrate a base of volcanic rock, the test was inconclusive according to the criteria set up, and the many scientists who had had a part in studying the core and the marine life on the outer slopes of the atoll were forbidden to express in the official report[2] any opinion of the bearing of the results on the controversy. It is known, however, that the distinguished specialists,

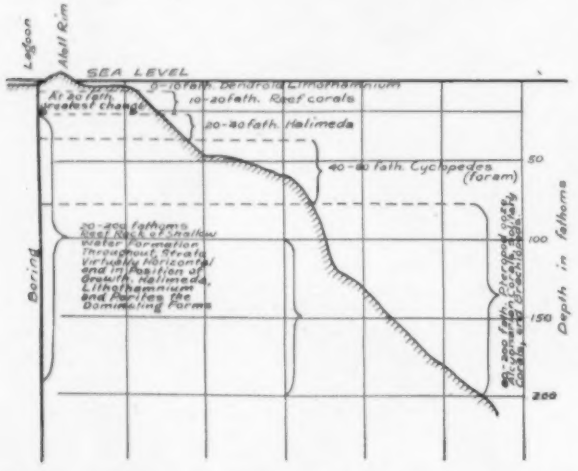


FIG. 2. Diagram to illustrate the conditions found at Funafuti, according to Sir Edgeworth David.

mainly zoologists, were practically unanimous in the opinion that the results furnished a definite proof of the essential correctness of the Darwin-Dana theory.[3]

The reasons for this uniformity of view will be clear from the following results of the investigation:

1. To the bottom of the boring the organic remains encountered were all from organisms which are now living only at very shallow depths.

2. These organic remains were all found erect —quite undisturbed from their position of growth.

3. The core was throughout composed of dolomite (calcium magnesium carbonate), and this change from the original carbonate of lime (dolomitization) is recognized to be a near-surface oceanic process (fig. 2).

MARGINAL MOUNTAINS A CONSEQUENCE OF THE SETTLEMENT

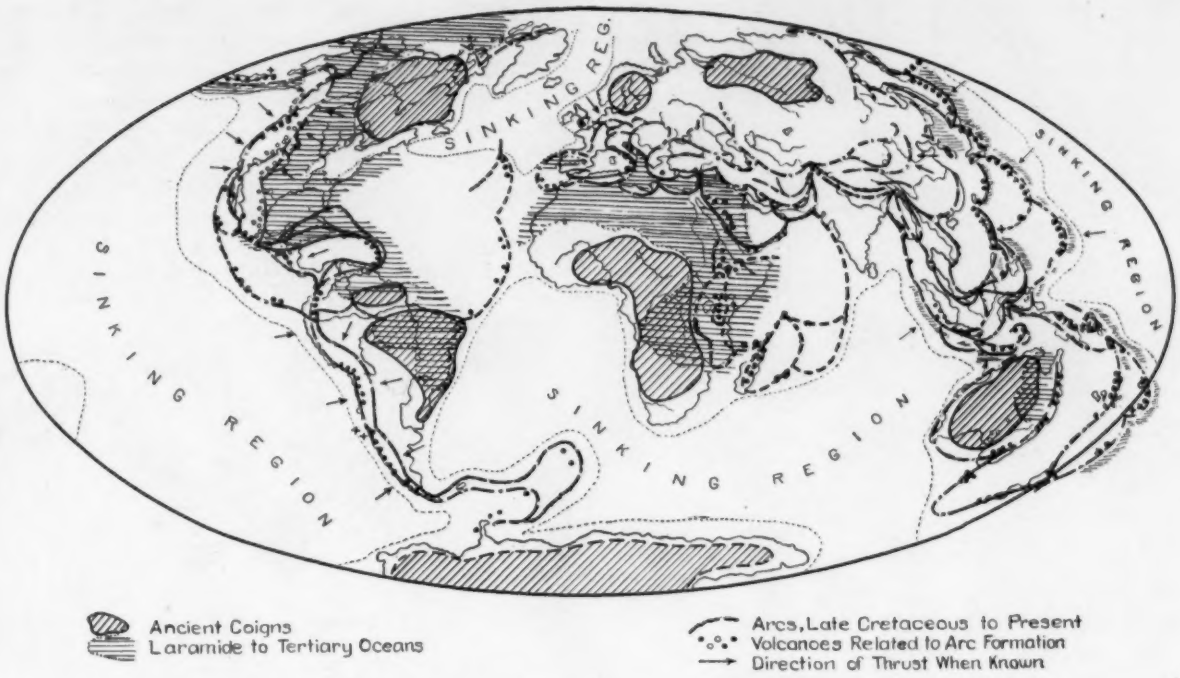
The settlement of the ocean floors and the resulting reduction of their superficial areas may be given, then, as the probable ultimate cause of the push or thrust outward against the continental margins and the wrinkles set up there in the earth's face. For the Pacific area the measure of the subsidence is in the neighborhood of 10,000 feet, or near 2 miles. The testimony of the reefs supports the view that this has taken place since the late Cretaceous age of geological time, and so it corresponds to the earth's latest period of mountain-building, that which raised the great mountain welts of Central Asia and the lofty Cordilleran backbone of the Americas (map 2).

Like the newest series marginal to the present Pacific Ocean, the earlier mountains partake of the same characters so far as mutilation permits these to be read—it is an arcuate pattern everywhere peripheral to the ancient coigns at the hearts of the continental areas. The ranges nearest to the central coigns are shown through their mutilations by erosion to be the most ancient. From them there is a gradation in age outward until at the outermost marginal zone mountains are even now rising from the sea to the accompaniment of strong earthquake joltings and of catastrophic outpourings of volcanic lava.

The inner and older arcs are today represented only by their roots, mainly the plutonic crystalline rocks which consolidated slowly at depths far below the vents to which their material ascended to issue as lava. To the devastating earthquakes which marked their youthful history there has here succeeded the placid stage of senility. Growth has long since ceased, and destructive agencies have largely accomplished their work.

[2] *Report of the Coral Reef Committee of the Royal Society. The Atoll of Funafuti, Borings into a Coral Reef and the Results.* London: 1–428, text and maps, 1904.

[3] Such a statement was made publicly by Sir Edgeworth David to the Second Pan-Pacific Congress held in Australia in 1923 and confirmed to me by another of the experts connected with the study. With the core open for inspection, a symposium on coral reefs was included in the program led by Sir Edgeworth David, who had carried the boring down to its greatest depth. (See also E. S. Skeats, *Amer. Jour. Sci.* **45**: 82, 1918.)



MAP 2. The salient features in the earth's face—the mountains raised during the latest of the earth's mountain-building periods.

2. THE UNDERTHRUST OUTWARD FROM OCEANIC SETTLEMENT

UNDERTHRUST LANDWARD AS AGAINST OVERTHRUST SEAWARD

The erection of a fold or wrinkle within the outer shell of the earth, since it is brought about by unbalanced forces, may be *a priori* considered in either one of two ways. The ever accumulating active stress eventually is no longer balanced by a passive stress set up within the rock material. The static condition is thereupon transformed into a dynamic one and a fold is initiated. Such a fold could be formed either (1) by the active force directed from behind and above the fold, *overthrust*, or (2) by the active force directed from the front and below, *underthrust*. These contrasted views are represented in figure 3.

The first of these conceptions as a cause of the erection of mountain arcs is the older view and was for a long time the more widely accepted one. It was originally developed by the great Viennese geologist, Eduard Suess, first in 1875[4] and later in his famous masterpiece, *Das Antlitz der Erde*, issued in four massive volumes.[5]

The opposing view, that of underthrust from in front of the arc, was first advocated by Willis[6] and later by the author.[7]

It must be clear that the Suess view requires *a priori* a condition of tension in the interior of the continents unless material is brought in either laterally or from below during the growth of

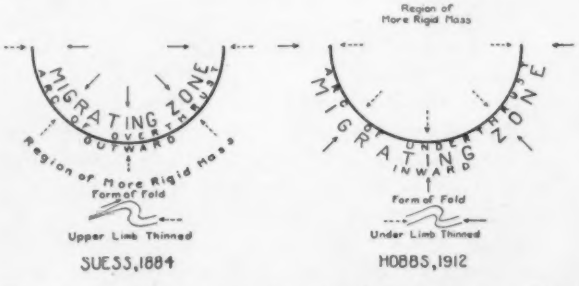


FIG. 3. Contrast of opposing views concerning the origin of arcuate mountains.

[4] *Die Entstehung der Alpen.* Vienna, 1875.

[5] Published at Prague and Leipzig between 1885 and 1909 and translated into French, English, and in part into Italian.

[6] Willis, Bailey. Mechanics of Appalachian structure. *U. S. Geol. Survey, Ann. Rept.* 13, pt. 2, 1893.

——. *Research in China* 2, *systematic geology.* Carnegie Inst. Washington, 1907.

[7] Hobbs, W. H. *Earth features and their meaning:* 436–438. New York, 1912.

——. Mechanics of formation of arcuate mountains. *Jour. Geology* 22: 71–90, 166–188, 193–208, 1914.

——. *Earth evolution and its facial expression,* chs. x and xi. New York, 1921.

mountains. When late in his study Suess came to treat the great Cordilleran mountain system of the Americas, his assumption of overthrust from the back encountered great difficulties which he could not harmonize with his theory.

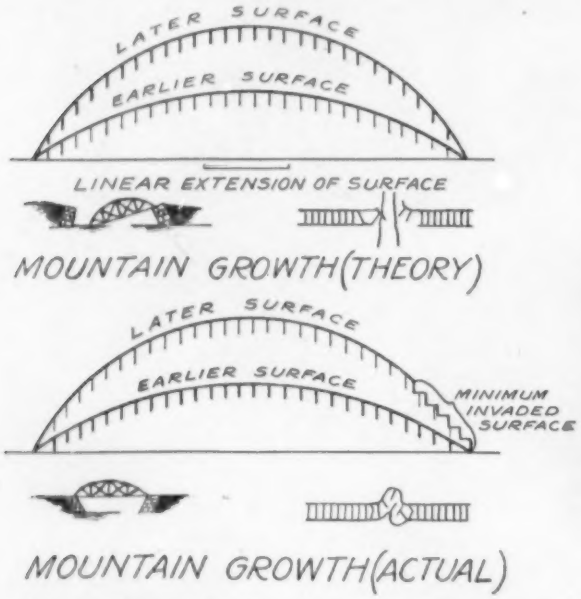


FIG. 4. Diagrams to illustrate the contrasted effects of tension (upper diagram) and compression (lower diagram) when present in rising mountain arcs. Below each is illustrated the effect on bridges and on railroads for each at the time of an earthquake.

The folds of the Andes were found to face outward on *both* flanks with the included area much circumscribed (map 1). He therefore felt obliged to consider the great Cordilleran system as an exceptional case which he designated the "Andean Structure."

Moreover, inasmuch as the process of folding augments rigidity, the Suess view of overthrust requires the outer arcs next the sea to be the more ancient and the others in succession more youthful, with the youngest of all next the coigns. Erosion effects and the seismicity unite in proclaiming loudly that the reverse is the fact. If the Suess view were correct, then in closely compressed folds the upper limb should be the thinner one, though the reverse is the rule (fig. 3, below).

COMPRESSION OF A MOUNTAIN ARC DURING ITS ERECTION

It would be natural to assume that mountain arcs in the process of erection should be in a state of tension for the reason that the anticlinal arch is thereby lengthened (see upper diagram in fig. 4). As a consequence the joints and any other openings of the rocks within the arch should in such case tend to open, and any continuous linear structures which traverse the arc should either be stretched or be torn apart with the rupture surfaces separated by a gap. Railroads and bridges, curbs, water and sewer pipes, and telegraph and telephone lines should all alike reveal this condition of tension. A study of such structures reveals, however, that without exception these structures have all been shortened at the time of earthquakes, and this whether they ran transverse to or along the arc.[8] Of hundreds of

FIG. 5. Buckling of rails on the approach to the Kisagawa railroad bridge after the great Japan earthquake of 1891.

FIG. 6. The Nagaragawa bridge after the great Japan earthquake of 1891. Though dropped into the river and given a sinuous course, all spans were still connected and the rails of the approaches indicated compression. (After Milne and Burton.)

[8] Hobbs, W. H. A study of the damage to bridges during earthquakes. *Jour. Geology* **16**: 636–653, 1908.
——, *Earthquakes, an introduction to seismic geology:* 183, 228–231, pls. XIV and XXII. New York, 1907.

examples examined from many earthquakes, there were no exceptions.

This apparent paradox is explained when the seaward margins of the arcs are examined, for a very considerable invasion of the area of the arc from the sea floor is everywhere indicated. On all coasts of the Pacific area, within which, as we have seen, a large subsidence has taken place, coastal staircases made up of successive strands are found on the seaward side of the arc. These represent elevations measured in hundreds or even thousands of feet and invasions of the arc which may run inland for tens of miles (figs. 4 and 7).

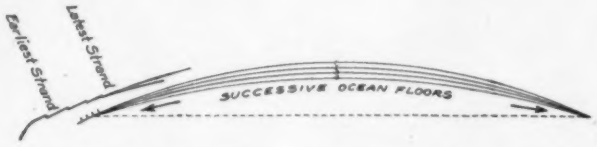


FIG. 7. Diagram to illustrate the relation of a progressive settlement of the ocean floor sector to the elevated Pacific strands.

Within these coastal staircases the "risers" are sea cliffs cut by the waves, and each represents the measure of uplift during a single strong earthquake or a series of closely related quakes.[9] The "treads" are many times as broad as the cliffs are high, and these measure the individual invasions from the sea floor to correspond to the earthquake or earthquake series which elevated the cliff at its back.

It would appear that at the time of each uplift the joints in the rock tend at first to open and suck down surface water, but are almost immediately jammed even closer together as the thrust comes in from the sea floor. It is characteristic of earthquakes that underground water is expelled at the surface in newly formed springs and from "earthquake fountains" which operate for hours and flood the country.

RESOLUTION OF THE UNDERTHRUST AT THE BOTTOM OF THE CONTINENTAL SLOPE

If we now consider once more the mechanics of the process when underthrust from the sea floor encounters the upward deviation along the continental slope, this must bring about a resolution of the thrust into a component thrust along the surface and one which continues the original near-horizontal direction (fig. 8). The former tends to lift the beds into an anticline, the latter to compress the arch when raised. The initial outward inclination of the surface which here starts the erection of the anticline was long ago

FIG. 8. Resolution of the underthrust from the sea floor where it encounters the continental slope.

given the name *initial dip* by Willis.[10] Successive stages in the process of erection of the anticline are illustrated in figure 8.

3. THE PATTERN OF A SYSTEM OF MOUNTAIN ARCS

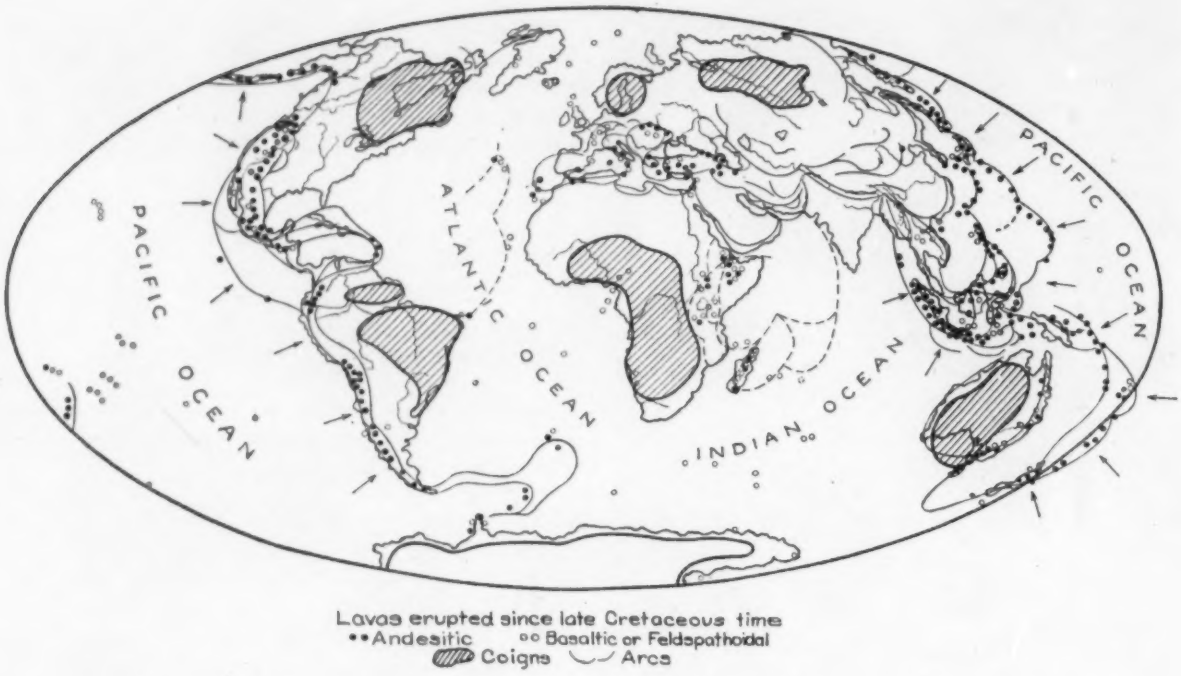
THE RELATIVELY RIGID COIGNS WITHIN THE CENTRAL AREAS OF THE CONTINENTS

Upon the continents the most ancient rock formations of our planet occupy central positions with the mountains arranged in peripheral zones about them (map 3). These early platforms were by Suess designated as coigns. They consist essentially of Pre-Cambrian rocks, about which through the ages of later geological time the continents have grown to their present dimensions through peripheral enlargement. From two earlier mountain-building periods, those at the end of the late Pre-Cambrian and the Permo-Carboniferous, the mutilated mountain remnants do not permit us to restore the features except very imperfectly.

In the latest period of the mountains (map 3), the older series outlines the Tertiary ocean distribution, while the later series outlines the present oceans. The mountains of the latest development display a pattern of outwardly (seaward) convex arcs joined in cusps, as do garlands which are hung upon a wall (map 3). In positions where frontal protection has been afforded by a shielding coign (Australia), thrusts from two opposing directions have resulted in great lateral compression of the arcs (Netherlands East Indies). Successive chains in a single arc may be numerous as in the Seistan arc of Baluchistan (map 4).

[9] Hobbs, W. H. The rate of movement in vertical earth adjustments connected with the growth of mountains. *Proc. Am. Philos. Soc.* **62**: 63–73, 1923.

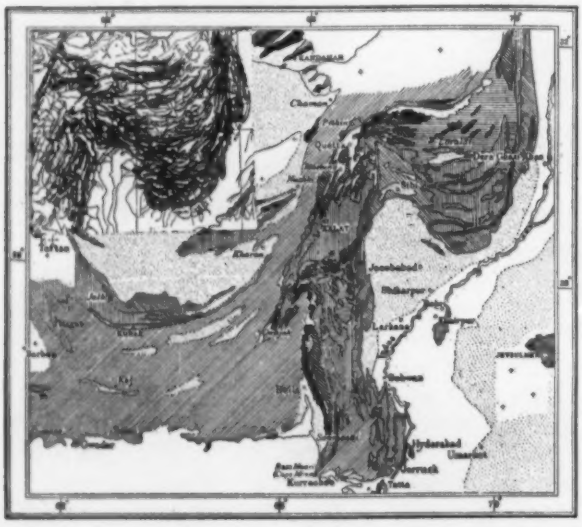
[10] Willis, Bailey. Mechanics of Appalachian structure. *U. S. Geol. Survey, Ann. Rept.* **13**, pt. 2: 246 ff., 1893.



MAP 3. The central coigns of the continents in their relation to the mountain arcs of the latest great mountain-building era.

THE LATEST ARC SYSTEM BORDERS THE PACIFIC OCEAN

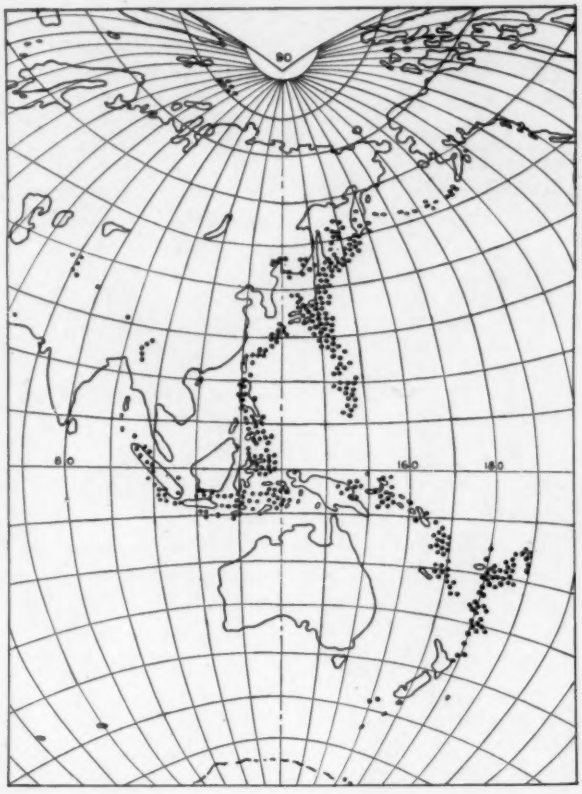
It is profitable to contrast the mountains which border the Pacific Ocean with those near the coasts of the Atlantic. The latter were formed in an earlier mountain-building period when the coast was in places in a different position from the present one. The infrequent earthquakes of the region indicate that mountain growth has ceased, and in those places where the former coast was in correspondence with the present one, the coastal staircase has been so mutilated by erosion as to be recognized only with difficulty. The Pacific border is by contrast the zone of greatest seismicity on our planet (map 5), and its coastal terraces betray only slight traces of erosional mutilation (fig. 9).



MAP 4. Successive mountain chains of the Seistan arc in Baluchistan. The mountain welts as features appear in the inset map, the flexures in the sedimentary formations in the main map. (After Vredenburg.)

From the map which exhibits the surface distribution of seismicity (map 5), it appears that earth adjustment is today at a maximum on the Pacific Ocean within two well-defined zones. These are on its western border from Japan southward to New Zealand, and on the opposite side of the ocean in the coast ranges of South America from Peru southward into Chile. Between these opposed borders of the Pacific Ocean lies its broadest expanse, and this includes nearly 180° of arc, as well as the galaxy of volcanic islands and atolls of Micronesia, Melanesia, and Polynesia—our earth's greatest region of crustal subsidence.

On the side of mechanics the arcuate form of the mountain ranges, always convex seaward, is



MAP 5. The earth's present-day seismicity of a part of our planet as indicated by teleseismic observations. The dots indicate intense seismicity ("earthquakeness"). (After Gutenberg and Richter.)

explicable on the basis of the directions of underthrust, which arrives in a broad converging fan of azimuths.

The earthquakes, which are the first manifestations to us of the jolting uplifts that produce the mountains on the rim of the Pacific, occur each within a segment which is comparable in extent to those of the arcs themselves. In this may lie the reason for the segmentation itself. Each greater earthquake within the Pacific's rim records the time when within that particular segment the stresses had accumulated sufficiently to overcome the resistance to adjustment inherent in the material. When this has been overcome, the potential energy is suddenly transformed into actual or dynamical energy and adjustment ensues. The stress condition is thus more or less completely relieved—completely for a central area and less so for sections increasingly distant from the central area.

Portions of the Pacific mountain rim most remote from that already relieved at the time of the earthquake are the ones most likely to find relief by the next succeeding earthquakes. Still later the probability of an early earthquake will be greatest for segments of the rim which lie midway between those already relieved of stress by definite earthquakes.

Such a sequence of successive great earthquakes occurred within the Pacific rim in the period 1906 to 1923, and better than any other it illustrates the segmentation of the zone of growing mountains.

For a considerable period the rim of the Pacific had been immune from a great earthquake when on January 29, 1906, one took place in Colombia, South America. A month and a half later (March 17) occurred the great earthquake in Formosa on the opposite side of the Pacific rim. Only a month later came the devastating California earthquake (April 15) about two-fifths the way from Colombia to Formosa. Dr. F. Omori, the great Japanese seismologist who had come from his distant home to study the San Francisco area of destruction, on his departure for Japan forecast in a press interview that the next great earthquake in the series would take place in Chile. Before his ship had arrived in Yokohama there occurred in Chile (August 17) the great Valparaiso earthquake, and on the same day but some hours later one in the Aleutian Islands.[11] These earthquake segments were each midway in portions of the Pacific rim then unrelieved. Dr. A. C. Lawson then forecast [12] that the next great earth-

[11] *Bull. Japanese Earthquake Committee* **1** (1): 23, 1907.

[12] In a lecture delivered in March, 1907.

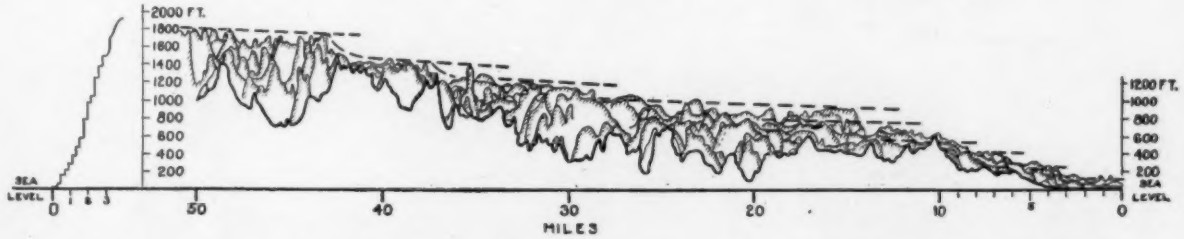


FIG. 9. Comparisons of raised terraces, Pacific and Atlantic. (From data by Tangier-Smith and Barrell.)

MAP 6. The successive segmental uplifts on the mountain rim of the Pacific within the period 1906–1931. (After Hobbs, 1909, except for segments VII–IX.)

quake would take place in Mexico, since there was there to be found the longest still unrelieved segment of the Pacific rim. His forecast was promptly verified on April 18 by the Guerrero earthquake. In April, 1909, I prophesied that the next great earthquake of the series should occur either in the Japan-Kamchatka segment or in that of New Zealand, since neither had yet been relieved.[13] This prognostication was doubly verified by the great Tokyo earthquake of September 1, 1923, and by the Napier, New Zealand, earthquake of February 3, 1931 (map 6).

[13] Hobbs, W. H. The evolution and the outlook of seismic geology (read April 24, 1909). *Proc. Am. Philos. Soc.* **48** (192): 289–290, map, 1909. This was repeated in lectures and in a press interview in the spring of 1923.

If the initiation of the arcs composing the Pacific rim of mountains resembled in its segmentation the latest uplifts, the cause of the segmentation would seem to be found in the partial nature of the relief of stress at any one time. Once segmented, the reliefs at the time of successive later earthquakes would be likely to increase the curvature of the individual arcs.

4. MOUNTAIN ARCS ARE FLEXURES WITH A DOWNWARD-BENT PORTION (SYNCLINE) ON THE SEAWARD SIDE OF THE UPWARD-BENT PART (ANTICLINE)

IN EARLY STAGE FLEXURES ARE SYMMETRICAL, BUT LATER UNSYMMETRICAL WITH STEEPER LIMB ON THE SEAWARD SIDE

We have already seen that in the evolution of an anticline the profile is first nearly symmetrical, but soon becomes unsymmetrical with the steeper limb at the front. This lack of symmetry is increased with the continued compression of the anticline (fig. 8). Perhaps the best youthful arc to display the near symmetry is on the island of

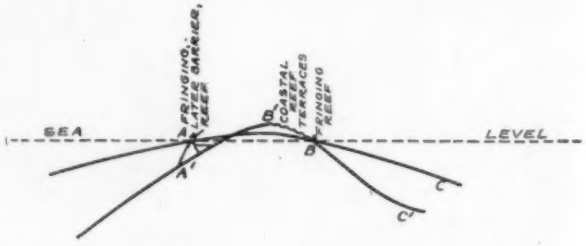
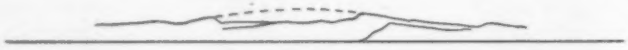


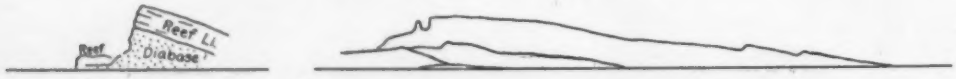
FIG. 11. Two early stages in the evolution of an island arc: *ABC*, early stage; *A'B'C'*, later stage.

Curaçao off the coast of Venezuela (fig. 10). As the anticline continues to rise, it is compressed. This develops terraces on the front and usually a barrier reef on the back (fig. 11). A collection of



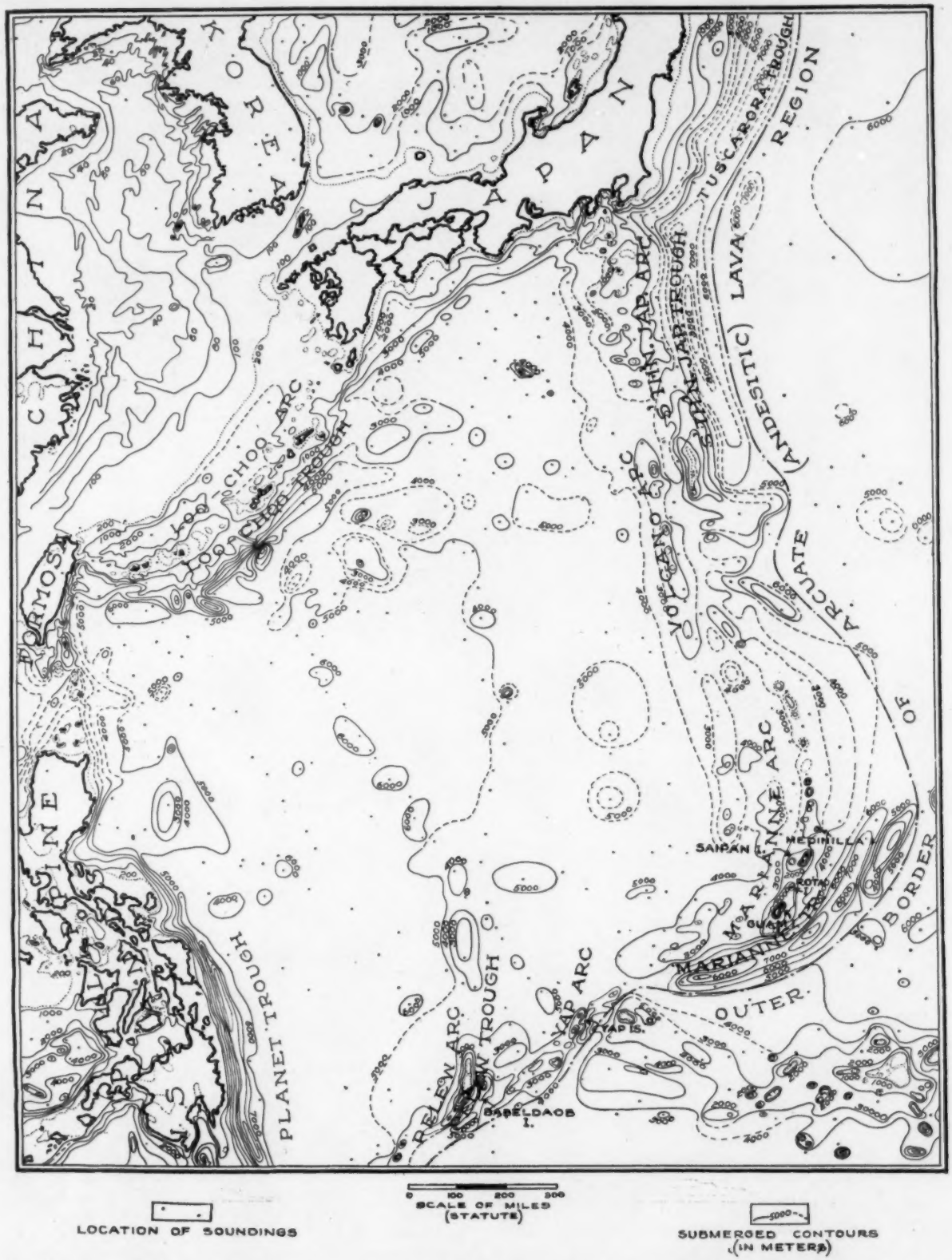
Looking northeast from harbor of Willemstad

South coast, looking west from off Willemstad



Below Fort Nassau, Schottegat — South coast, looking east from off Willemstad

FIG. 10. The nearly symmetrical anticline of Curaçao. (From sketches by the author.)



MAP 7. Map of the oceanic region between Japan and the Philippine Islands. (After Japanese hydrographic charts.)

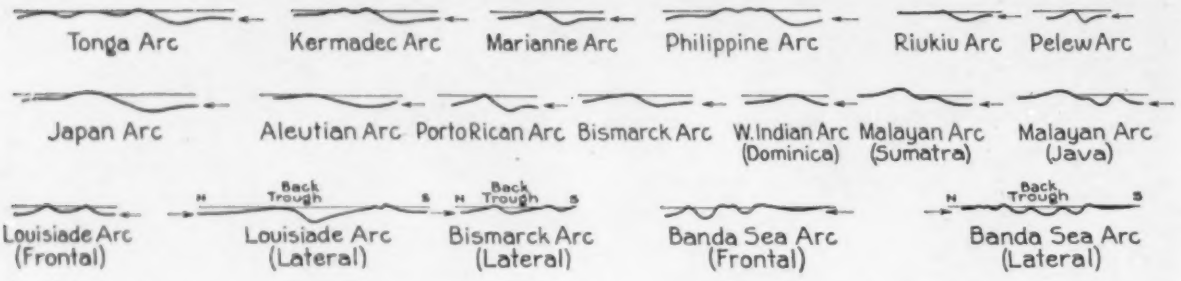


FIG. 12. Sections of rising arcs. Vertical to horizonal scales, 12: 1. → Direction of thrust.

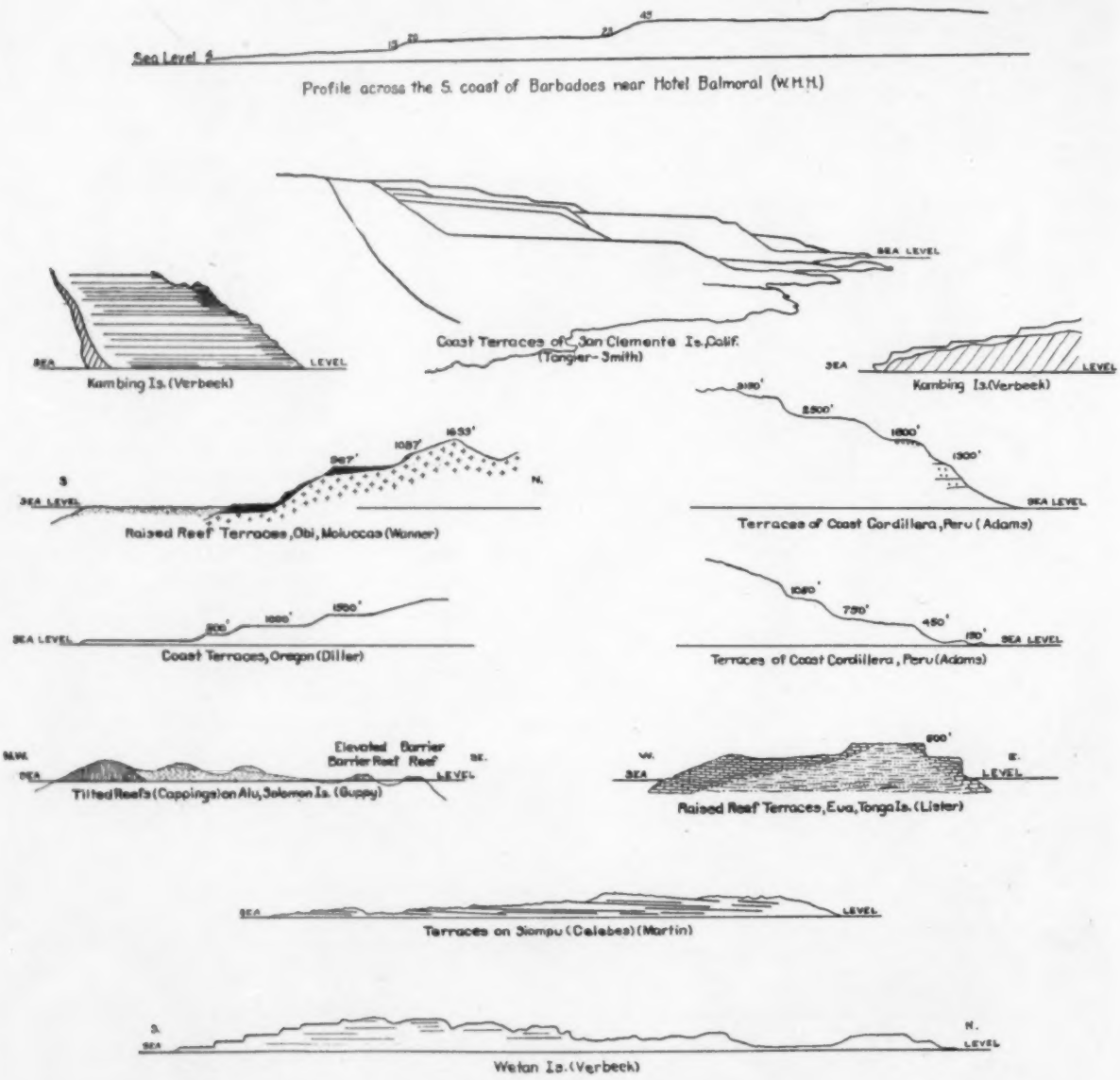

Sea Level
Profile across the S. coast of Barbadoes near Hotel Balmoral (W.H.H.)
SEA LEVEL
Kambing Is. (Verbeek)
Coast Terraces of San Clemente Is. Calif. (Tangier-Smith)
Raised Reef Terraces, Obi, Moluccas (Wanner)
Terraces of Coast Cordillera, Peru (Adams)
Coast Terraces, Oregon (Diller)
Terraces of Coast Cordillera, Peru (Adams)
Elevated Barrier Reef
Tilted Reefs (Coppinger) on Alu, Solomon Is. (Guppy)
Raised Reef Terraces, Eua, Tonga Is. (Lister)
Terraces on Siompu (Celebes) (Martin)
Wetan Is. (Verbeek)

FIG. 13. Terraced coasts of recent uplift about the Pacific Ocean; Barbadoes in the figure should be Barbados.

profiles of early stages in the evolution of mountain arcs is included in figure 12.

Off the coast of Asia the pattern of the rising mountain arcs within the area to the southward of Japan is a relatively simple one of very flat curvature, always the indication of an early stage. In fact, nowhere else in the world can we find so many arcs in the stage of early youth. Here also soundings of the sea recently made by the echo method have come to our aid by revealing the features of the ocean floor to supplement those of the islands (map 7). In these youthful arcs the terraces are generally found on both sides wherever the arc has risen from the sea floor, and they may extend up to an elevation of a thousand feet or more (fig. 13). However, in proportion to the time which has elapsed since their elevation they have been modified by erosional agencies.

MAP 8. The position of the active volcanoes relative to the anticlinal arcs off the Asiatic coast.

5. ACTIVE VOLCANOES ACCOMPANY THE ERECTION OF ARCS

VOLCANIC VENTS APPEAR ON THE BACK OF THE ARC

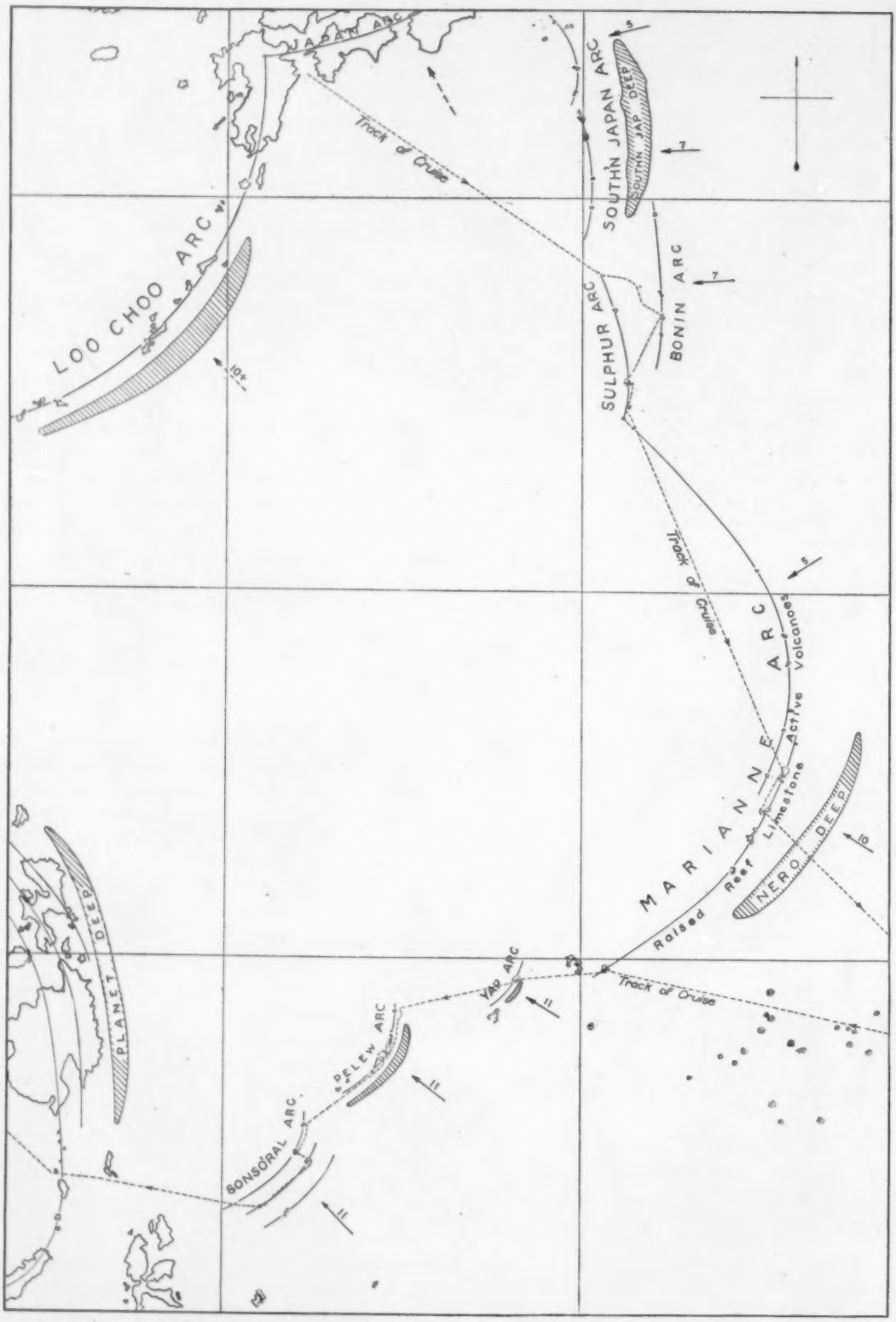
All mountain arcs of youthful stage are characterized by a line of active volcanoes which appear on the back (landward) side of the anticlinal arc (map 8).

The volcanoes have usually made their appearance before the anticline has risen above the waves. The Marianne arc supplies an illustration (map 9), as does the Iwo (Sulphur or Volcano) arc. Back on the continents, where erosion has removed the upper parts of the anticlines, most former surface features have disappeared and the volcanoes are no longer in evidence. In their place we find their roots, either the consolidated and generally crystalline rock that fills the pocket from which the lava was supplied, or that filling the pipe through which it ascended to the vent (map 10). In much-degraded mountain ranges throughout the world the maculae or reservoirs which were the source of the lava of an earlier stage are now represented by granitic cores (map 11).

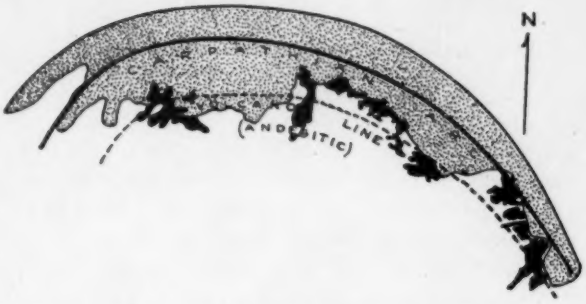
INITIAL LAVA EFFUSIONS HAVE THE COMPOSITION OF MUDSTONE—ANDESITE

The initial and early effusions from the volcanoes on the backs of mountain arcs betray but little range in chemical composition. Though lavas from all parts of the world vary in silica content from more than 80 percent to less than 40 percent, those exuded from the volcanoes of the mountain arcs fall near the average—from 60 to 65 percent of silica content. They thus correspond roughly in composition to a mud rock (shale or slate), and they are described as andesites. This is close to the average composition of rocks of all kinds, which is about 59 percent silica.

The later effusions from the vents on mountain arcs show, however, no such constancy of chemical composition, but they are as a rule much poorer in silica content (basalts), though sometimes richer (rhyolites) and often also richer in the alkali constituents (lenads, which contain as an essential constituent the minerals nephelene or leucite). Variety in composition, rather than constancy, is characteristic of the later extrusions. Rather generally the later extrusions issue from a later formed line of volcanoes farther back on the anticline (fig. 14) or even on a secondary anticline developed at the rear.



MAP 9. Outline map of the young arcs off the Asiatic coast between Japan and the Philippines.



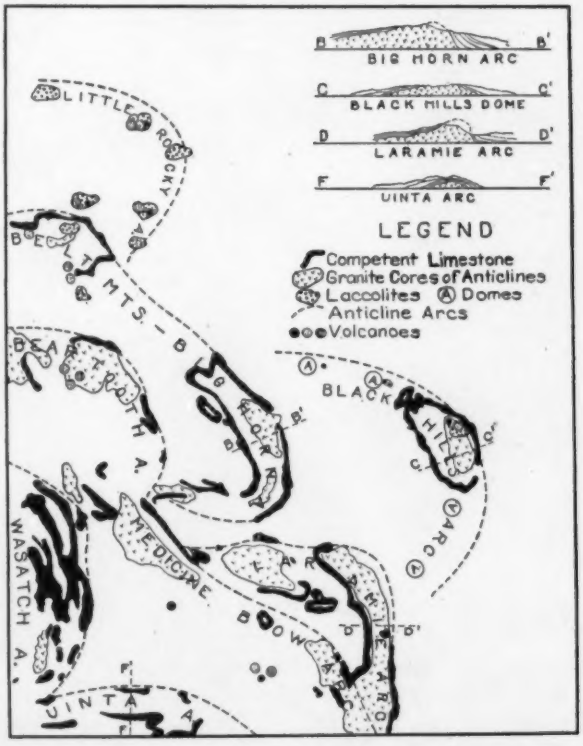
MAP 10. The Carpathian arc in southeastern Europe, with lava shown in black at the back of the anticline.

6. THEORY OF ORIGIN OF THE VOLCANIC EFFUSIONS

FUSION OF ROCK BENEATH A COMPETENT ARCH

In the opening section of this paper the reasons have been passed in review for the belief that the earth's interior is rigid and, in the near-surface shells at least, in a solid state, even though its temperature at depths of only a few miles is above the melting point of the rock.[14] The rock is kept in an unfused state only because of the superincumbent load, the weight of the rocks above. If a part only of this load were to be lifted off at any place, the rock beneath would at once become fused and produce a local pocket or *macula* of lava material, *magma*.

[14] For a somewhat fuller discussion, see Hobbs, *Earth evolution and its facial expression*, chs. ii and iii, 1921.



MAP 11. Denuded arcs off the front of the Rocky Mountains, showing, beneath a competent anticline of limestone, cores of granitic rock.

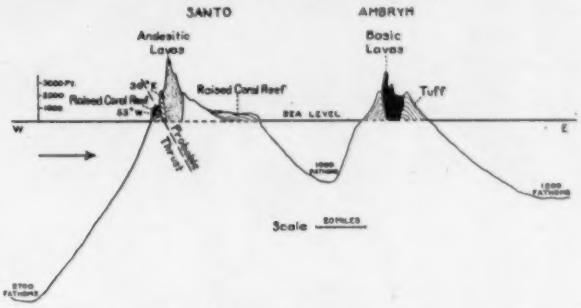


FIG. 14. Section across arc of New Hebrides. (After Mawson.)

There are two quite different geological processes known which seem competent to lift the load locally, and one of these is the formation of an anticline or upward bend in the rock strata such as forms in the initial stage of the growth of a mountain arc. An anticline is an arch, and if of strong rock such as limestone, it can sustain load as does the arch of a bridge.

Of the three main types of bedded or stratified rocks (sandstone, mudstone or shale, and limestone) formed as marine sediments, limestone is much the strongest, and the order of deposition is usually such that it lies over shale. When now an anticline is raised, it forms what is called a competent arch which lifts the load at least in part from the shale beneath. The shale (mudstone) thus fuses to produce a lava of intermediate silica content, andesite (fig. 15).

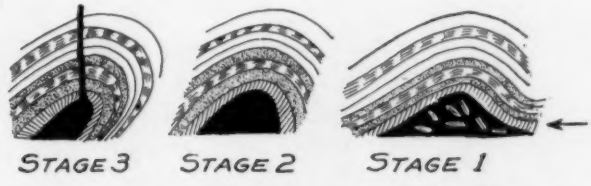


FIG. 15. Successive stages in the fusion of mudstone beneath a competent arch of limestone to produce an andesitic lava. The continued compression of the anticline forces the lava to the surface where it issues in a vent at the back of the arc.

In the laying down of marine sediments during a transgression of the sea over the land, the sandstones laid down near shore, and therefore at the bottom of the series, come to be gradually over-

laid by intermediate siliceous mudstones, then by pure mudstones, and from greater sea depths calcareous mudstones, and, lastly, pure limestones to form the top of the series. The pure mudstones are not only the weakest of the entire series, but they are also the most fusible. They should thus be the first to fuse under the competent arch of limestone and calcareous shale (fig. 16, at top).

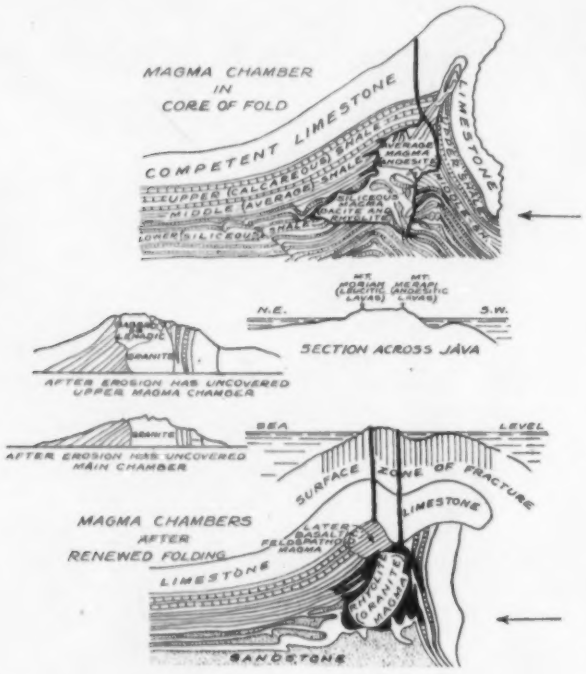


FIG. 16. Profiles to illustrate the possible fusion of shale beneath a competent arch of limestone in the raising of the anticline in a mountain arc (top). The later development of more siliceous and less siliceous lava types is also suggested (bottom). The uncovering of the lava cores by erosion is also illustrated (middle).

LAVA RISES TO THE SURFACE THROUGH THE CONTINUED COMPRESSION OF THE ANTICLINE

A continuation of the evolution of the arc and the resulting compression of the anticline acts upon the magma pocket as does a syringe held in the palm upon the included liquid. The lava is expelled from the pocket along a vertical conduit to the surface, aided by any available fissure in the limestone. Because of the unsymmetrical form of the anticline with steeper limb at the front, the volcanic vent will always be located at the back of the arc (figs. 14–16).

7. EVOLUTION IN THE PLAN OF THE ARC

EARLIER STAGES OF RELATIVELY FLAT CURVATURE WITH TROUGH AT THE FRONT

Comparison of many different arcs that are in different stages of development indicates that the evolution is from the simple to the complex. This is particularly to be observed in the curvature of the arc as displayed in the plan.

A very young arc approaches a straight line—its radius of curvature is very long. This is illustrated by the first stage in figure 17, in which the arc subtends an angle of only 45°. At this stage the arc is represented above the sea by a line of active volcanoes, the anticline still submerged,

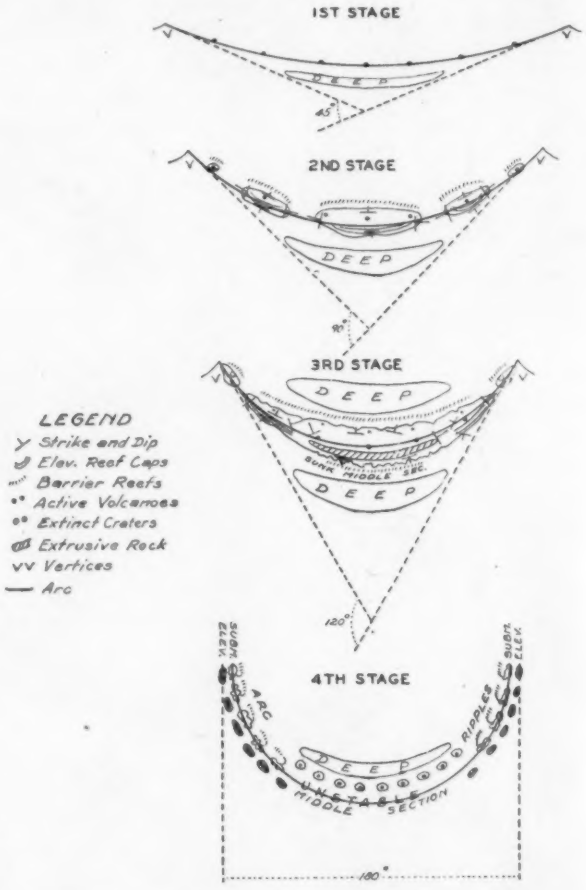
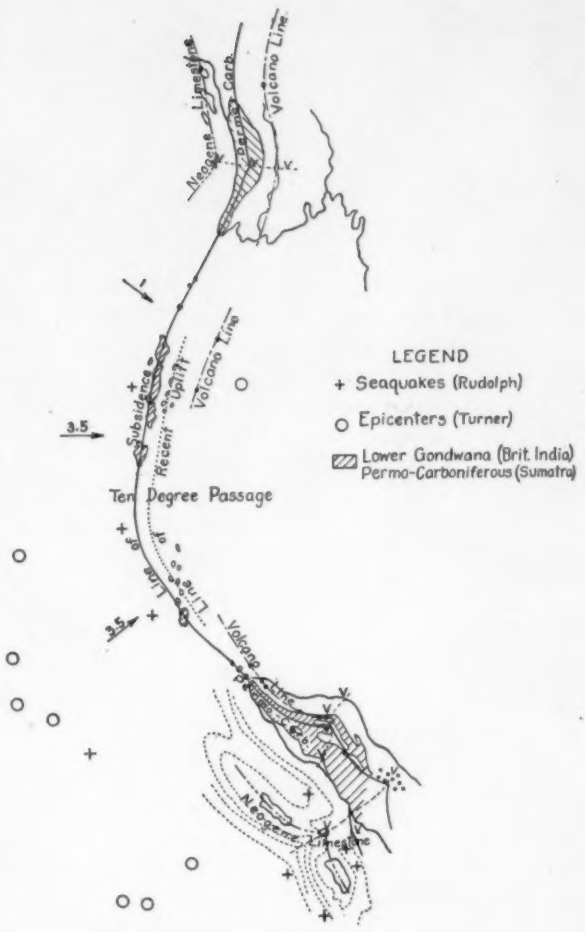


FIG. 17. Evolution in the plan of a mountain arc.

but a long and narrow trough is at the front. The Bonin, Iwo, and Marianne arcs, all to the south of the "elbow" of Japan, are examples (map 9).

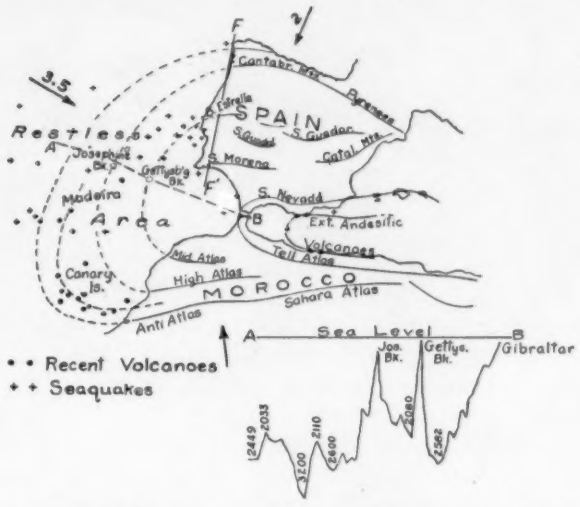
A second stage in the evolution is one in which the arc subtends a larger angle, where the anti-



MAP 12. Plan of the Andaman arc.

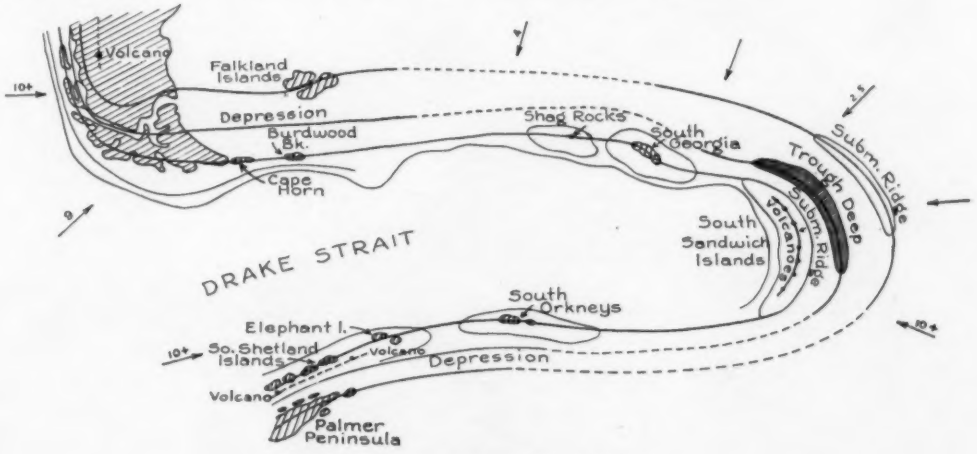
cline is represented above sea level by a series of islands composed of either marine or volcanic sediments (tuff and agglomerate), or both. The trough is, as a rule, shorter and broader. The islands are terraced, especially on the front, and there are barrier reefs at the back. A seismic zone is accentuated at the front. The Andaman arc which joins Sumatra to Burma is an example (map 12).

In a still later stage, here arbitrarily chosen as the third stage, the curvature of the arc is much sharper; the islands which represented the anti-

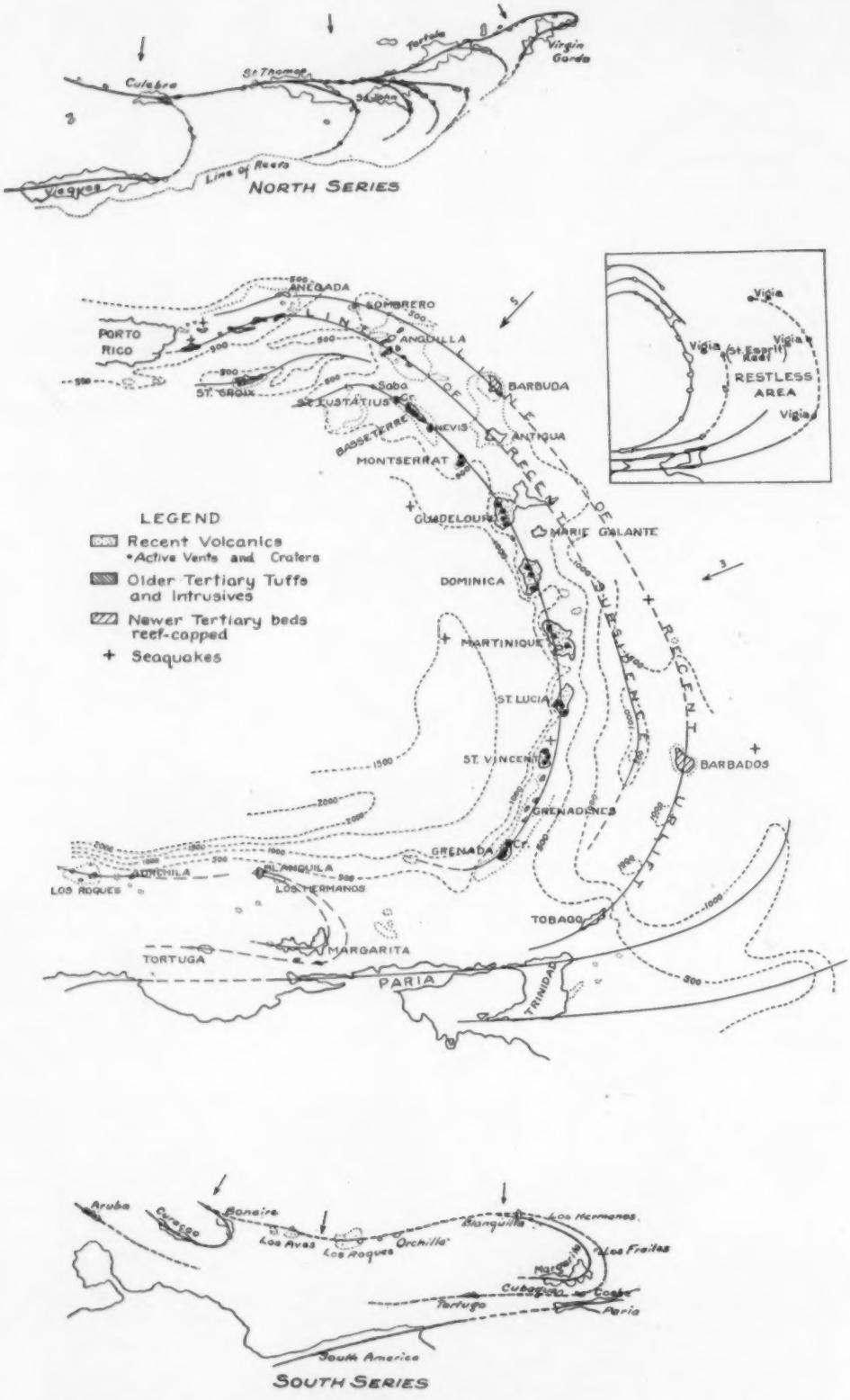


MAP 13. The compressed Iberian-Mauretanian arc.

cline in stage 2 and were the crests of its longitudinal wrinkles, have now been joined as the anticline emerged further. A new line of volcanoes has appeared behind the first one. The trough may be at the front or back or both, and the middle section of the anticline is very unstable. Off the front is a special seaquake zone, and there are records of many vigias—indicating islands which appear only to disappear, and perhaps later reappear.



MAP 14. The compressed Drake Strait arc.



MAP 15. The extremely complex Antillean arc, with elaborate ripples at the north and south and vigias off the front (inset).

In stage 4 is represented the extreme development of the arc, one which subtends an angle of 180° or more. Of this stage there are a good many examples. They appear to represent two different classes. In one the front of the arc has been protected by a coign or other land mass, while the sides have been shoved toward each other by subsidences in two different oceans on opposite sides. The outstanding example is the Sunda-Banda arc between Australia and the mainland of Asia (map 27). Another example is the Iberian-Mauretanian arc (map 13), which, though not protected from the front, was compressed between an earlier sea at the south and the North Atlantic. Here the entire front of the middle section has dropped down along a great fault and now represents a peculiarly restless area in the sea.[15]

Two noteworthy examples of the other type are supplied by the compressed arc of Drake Strait between extreme southern South America and the Palmer Peninsula of the Antarctic Continent (map 14). In this extraordinary arc, though there was a minor lateral thrust from the North Atlantic at the north, the maximum thrust probably came from the Pacific Ocean at the back of the arc. The extremely complex Antillean arc probably was developed in a similar way, with the maximum thrust from the Pacific Ocean subsidence at the back at a time before the elevation of the Isthmus of Panama had closed the gap (map 15). Here, as in the case of the Iberian-Mauretanian arc, the entire front has dropped into the sea and over it there is now a notably restless area with records of vigias and extreme sudden changes of level.

RISING ARCS THE LOCUS OF STRONG EARTHQUAKES

As we have already noted, the erection of mountain arcs takes place at a rate which, measured by prevalent views of geological adjustments, is startlingly rapid. This is attested by the strong earthquakes at the instants in which uplifts of the lands have amounted in some cases to many tens of feet and subsidences at the bottom of trough deeps have been measured in as many hundreds of feet. As shown by the large-area and small-scale maps which display the distribution of earthquakes, these zones are identical in position with rising arcs of today (map 5). The large-scale maps of smaller areas likewise attest this identity. The *latest* arc to rise in any area is sure to be characterized by the *strongest* earthquakes (maps 25, 27). How important earthquakes may be in indicating the areas of adjustment, has been shown in study of the compressed arcs brought under review in the preceding section.

PART II, MOUNTAIN ARCS OF THE SOUTHWEST PACIFIC REGION

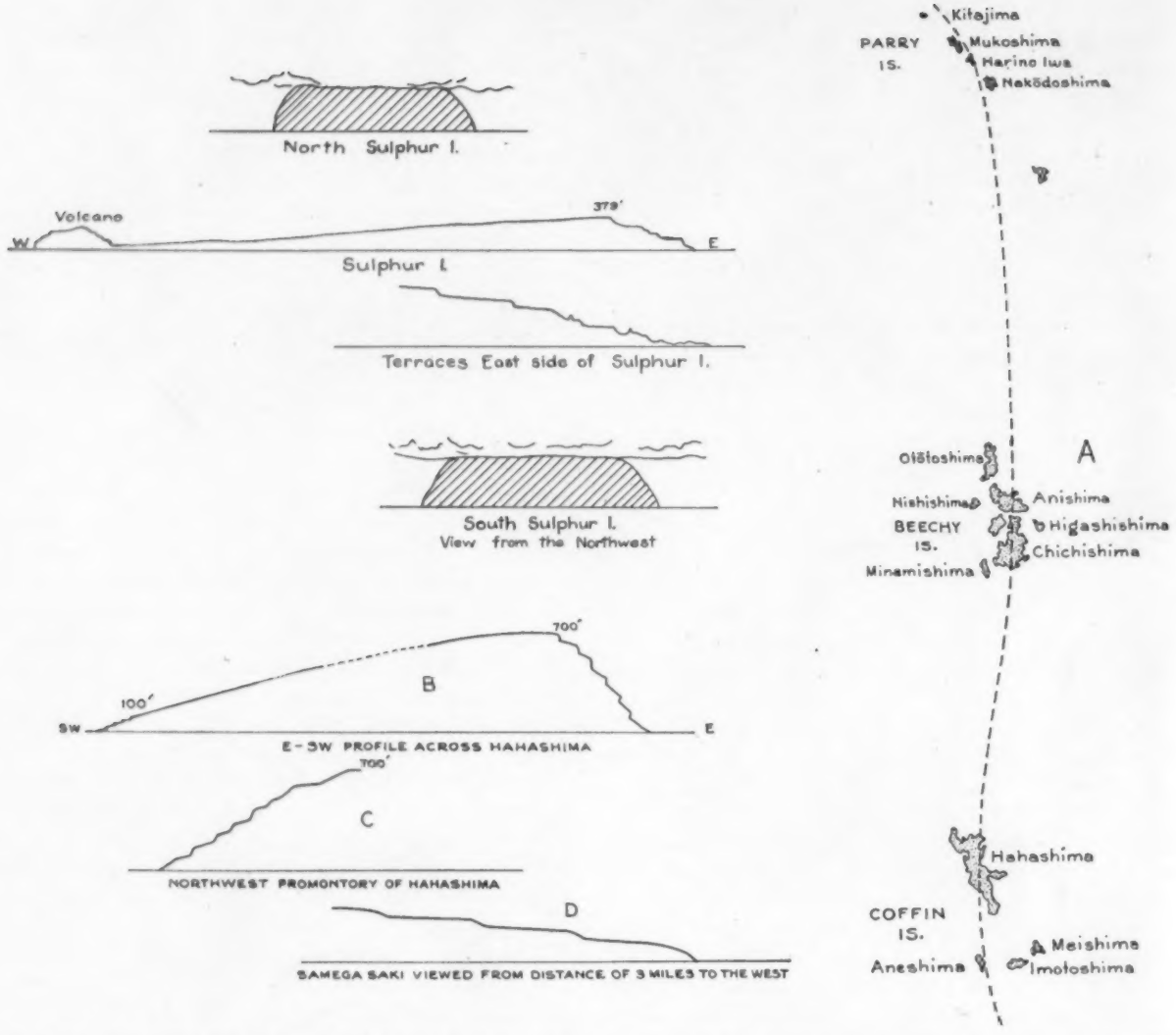
1. THE BONIN ARC

The Bonin (Ogasawara) arc, as shown at the right (*A*) in map 16, is not well outlined by the grouping of islands, and may indeed be in reality two arcs with a cusp at Hahashima, largest of the islands. All islands appear to be composed of andesite lava and andesitic tuff and agglomerate. Professor Wakemidzu, of the agricultural department of the Imperial University of Tokyo, has described the lava as augite andesite. There is a reef on the west coast of Chichishima, but none on the east coast. Specimens of the lava which I collected from near the port of Bonin, where it is found in pebbles (beach shingle) in an agglomerate together with madrepore and brain coral, were examined in thin section by Professor Walter F. Hunt, of the University of Michigan, and analyzed by Dr. R. K. McAlpine. It proves to be augite hypersthene andesite.[16]

The southernmost large island of the Coffin group is Hahashima (Mother Island). On its east coast, as reported to me by the director of the Museum at Bonin, is Sekimon (Gate of Star) promontory, which rises to about 700 feet above the sea and is terraced throughout, with the treads of the staircase dipping eastward as steeply as 50°. The promontory on the southwest (Okimuta) is only 100 feet high and, according to the same authority, has terraces with limestone cappings containing Tertiary fossils and with treads dipping 30° to the west. This would clearly indicate an anticlinal structure, and though I have often seen strongly tilted treads of this nature on arcs off the Asiatic coast, I have not found any as steep as 50°. I had no opportunity to land on the island. I sailed close to the northwestern promontory, which rises to about

[15] Hobbs, W. H. The unstable middle section of the island arcs. *Gedenkboek Verbeek, Verh. Geol.-Mijnb. Genootschap Nederland en Kolonien, Geol. Ser.*, **8**: 219–262, 4 pl., 19 fig., 1925.

[16] Hobbs, W. H., and Walter F. Hunt. Petrography of an area of the Western Pacific. *Proc. Third Pan-Pacific Sci. Cong., Tokyo*, **1926**: 711–715, 1926 (1927).



MAP 16. At upper left, profiles of the islands of the Iwo, Volcano, or Sulphur arc (from sketches by the author). At right (*A*) map of the Bonin arc (after a modern Japanese map). At lower left, profiles of Hahashima.

950 feet, with westerly dipping terrace treads up to about 700 feet (*C* in map 16). Samega Saki, seen from the north, was terraced on a relatively gentle slope (*D* in map 16).

According to Petersen,[18] the western coast of Chichishima is much cliffed with great stacks off shore, and this I was able to confirm.[17]

The departure of the line connecting the islands of the Bonin group from a perfect arc has raised the question of whether it is a single arc or two or more. There is a marked trough along its eastern front, but Japanese hydrographic charts generalize this into an exceptionally long trench extending nearly to the "elbow" of Japan and concave toward the east, a most abnormal feature. Kotô has clearly joined several arcs together in his monograph.[18] The islands are deserving of extensive study, which will be necessary in order to clear up the uncertainties.

2. THE IWO ARC

The Iwo, Sulphur, or Volcano arc is one of the most youthful to be found anywhere. Its position is south and west of the Bonin arc and be-

[17] Petersen, Johannes. Beiträge zur Petrographic von Sulphur Island, Peel Island, Hachijo und Mijakishima. *Proc. Third Pan-Pacific Sci. Cong., Tokyo*, **1926**: 24–54, pl. II, 1927.

[18] Kotô, B. The Rocky Mountain arcs in eastern Asia. *Jour. Faculty Sci. Imp. Univ. Tokyo* **3**, pt. 3, pl. VIII, 1931.

tween Japan and the Mariannes or Ladrones.[19] The three small islands which are comprised in this arc are in a north-south, nearly straight line with a trough on the seaward side (maps 9 and 16, upper left). The North and South Islands are "steep-to" lava domes, but the middle island (Sulphur Island) is made up of beds of volcanic tuff which rise to 379 feet and are inclined gently to the westward. Terraces are strongly developed on the higher east coast, and a small volcano still "smoking" is at the west end. We have here the usual indication of an anticline (fig. 10). The island has emerged so recently that there is as yet absolutely no indication of erosion, nor any appearance of reefs. When visited, both the North and the South Islands "had their hats on" (map 16), but they are obviously lava domes of augite andesite.[20]

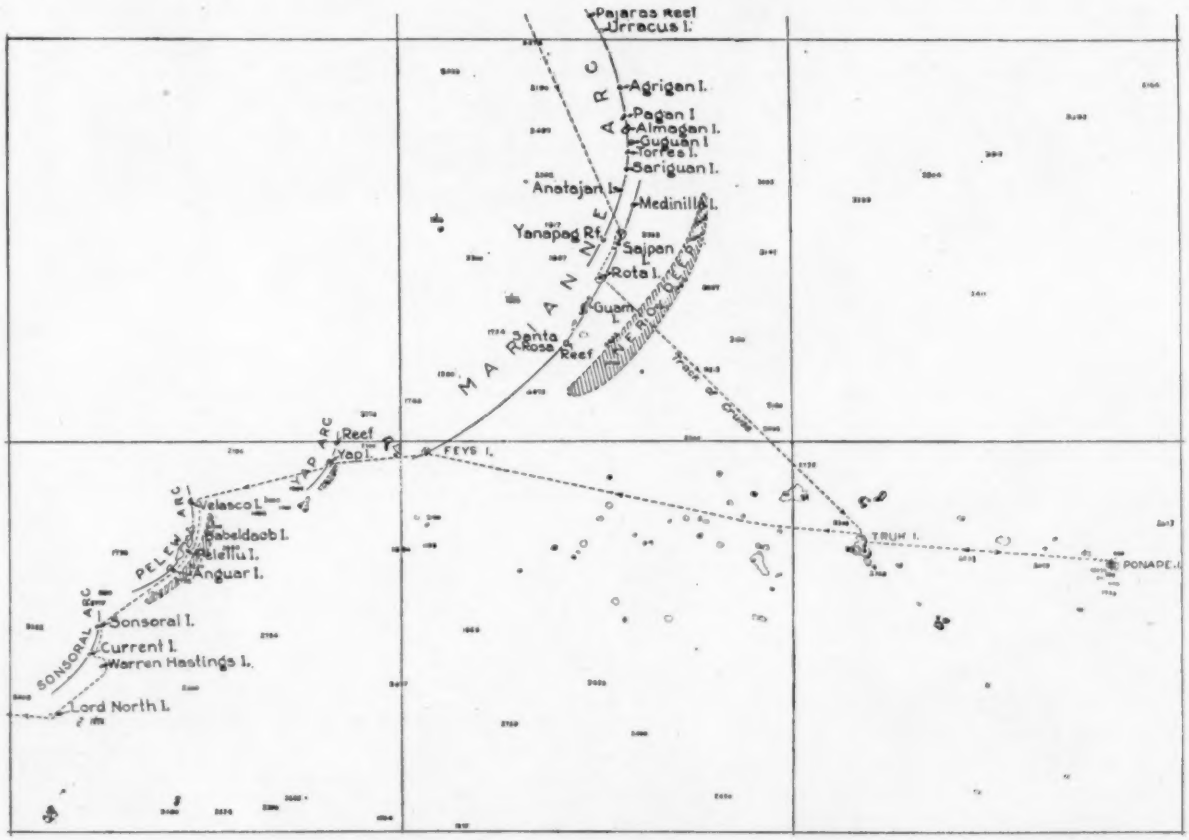
[19] The Iwo islands were practically without inhabitants, though a single canoe started out to our ship from the North Island. They were without any regular communication with the outside world and were visited by the *Matsumara Maru* in 1921 as the result of a destructive typhoon and the search for a Japanese liner which had not reported. War dispatches indicate great changes since 1921.

3. THE MARIANNE ARC

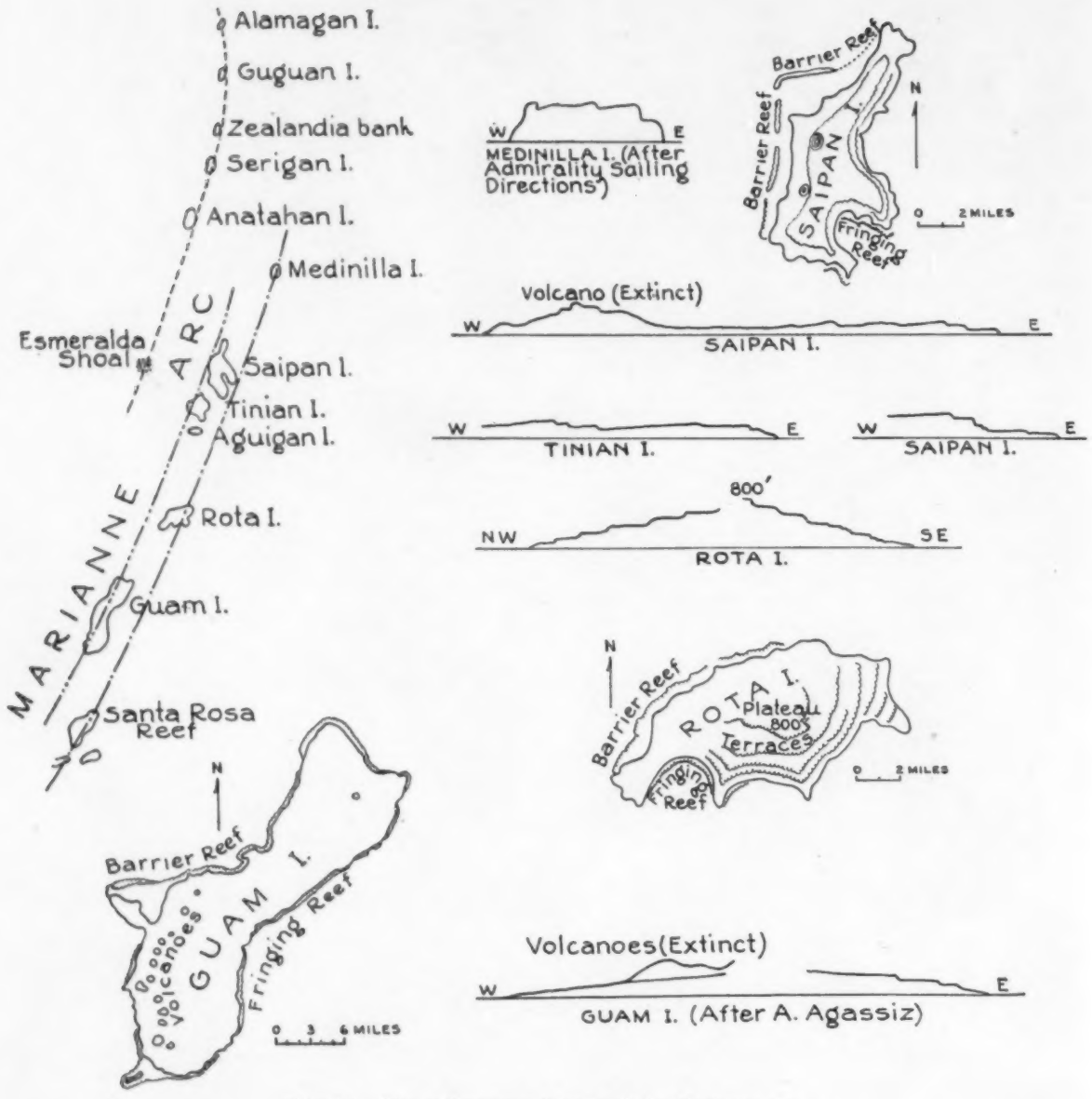
In contrast to the little-studied Bonin arc, which is somewhat difficult to outline, the Marianne (Ladrone) arc is a schematically perfect arc (map 17). It is about 700 miles in length and subtends an angle of about 45°. The southern section developed first, and its trough at the front is the profound Nero Deep, next to the Planet Deep the deepest in the world. Within this southern section of the arc are the islands of sediments, largely volcanic tuffs and reef cappings. In order from the north these are Medinilla, Saipan, Tinian, Aguigan, Rota, Guam, and in continuation southward Santa Rosa Reef, the raised atoll of Feys, and probably to be included Ululssi Reef.

Medinilla rises only about 50 feet above the sea and is terraced throughout. It is only about a mile in length, and on the southern and western

[20] Petersen, *op. cit.*



MAP 17. Arcs of the series from the Iwo arc on the north to the Sonsol arc at the south.



MAP 18. The sedimentary islands of the Marianne arc.

coasts are extensive sea caverns. Saipan is terraced and with reef cappings (map 18 and fig. 18). The beds slope gently westward, with extinct volcanoes on the west coast. Offshore there are barrier reefs. The lava of the volcanic effusions is hypersthene andesite.[21] One of the volcanoes has an elevation of 1,345 feet.

Tinian, like Saipan, is terraced throughout and without any considerable evidence of erosion except by the waves (fig. 18*D* and *C*). West of Tinian a distance of 17 miles is Esmeralda Shoal, and, as there are depths of 25 fathoms close to the reef, it is doubtless a volcano of the line at the back and in evidence mainly to the northward.

Aguigan Island, 42 miles northeast of Rota, is flat and terraced with high cliffs. Rota is made up of great terraces with reef cappings, the one at the top of great breadth at an elevation of 800 feet (map 18 and fig. 18*E'*). With little doubt this is now an air base of the first rank. At the southwest of the island is the diminutive boat

[21] Tsuboya, K., *Japanese Jour. Geol. and Geog.* 9: 208–210, 1932.

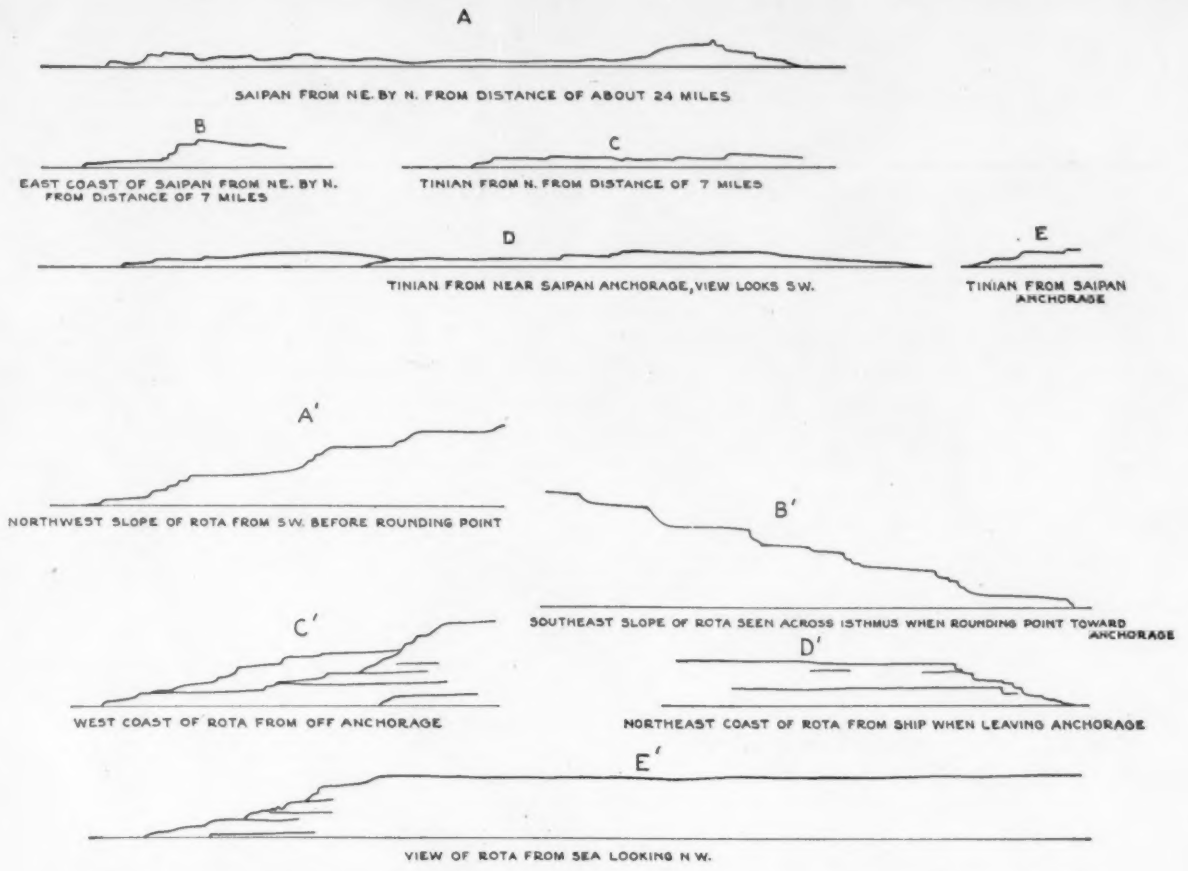
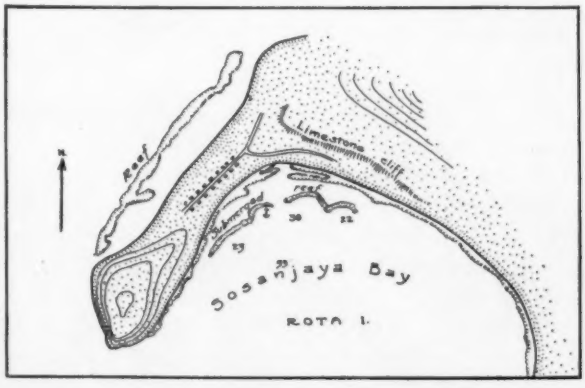

A
SAIPAN FROM N.E. BY N. FROM DISTANCE OF ABOUT 24 MILES
B
EAST COAST OF SAIPAN FROM NE. BY N. FROM DISTANCE OF 7 MILES
C
TINIAN FROM N. FROM DISTANCE OF 7 MILES
D
TINIAN FROM NEAR SAIPAN ANCHORAGE, VIEW LOOKS S.W.
E
TINIAN FROM SAIPAN ANCHORAGE
A'
NORTHWEST SLOPE OF ROTA FROM S.W. BEFORE ROUNDING POINT
B'
SOUTHEAST SLOPE OF ROTA SEEN ACROSS ISTHMUS WHEN ROUNDING POINT TOWARD ANCHORAGE
C'
WEST COAST OF ROTA FROM OFF ANCHORAGE
D'
NORTHEAST COAST OF ROTA FROM SHIP WHEN LEAVING ANCHORAGE
E'
VIEW OF ROTA FROM SEA LOOKING N.W.

FIG. 18. Profiles of the islands in the anticline of the Marianne arc. (From sketches by the author.)

harbor of Sosanjaya Bay behind a barrier reef, and another offshore reef is west of the bar which has joined a diminutive "Gibraltar" to the island (map 19).

Guam, the largest island of the arc, is similarly terraced, as may be seen from high points on Rota. At the front Guam has a fringing reef, and at the back an offshore barrier reef, a part of which encloses the naval base. On this side also are extinct volcanoes (map 18). The lava is augite andesite, a large part of its exposures covered by reef cappings. According to Agassiz,[22] the peak of Ata Rosa, east of Agana, rises 150–200 feet above the general level of the island, and Mount Mokana, another volcano, has an elevation of about 600 feet.



MAP 19. Detail of southwestern Rota.

Beyond Santa Rosa Reef the arc of the Mariannes is probably joined to the raised atoll of Feys and beyond to Ululssi Reef. In the *Chikuzen Maru* I was able to circumnavigate Feys and photograph and sketch its profiles (fig. 19).

The western line of generally active volcanoes of the Marianne arc lies back of the sedimentary islands with their extinct volcanoes. The volcanoes have reared their heads above the sea only at the north beyond the front line of the arc.

[22] Agassiz, A. The coral reefs of the tropical Pacific. *Mem. Mus. Comp. Zoöl.* **28**: 367–375, pl. 195, figs. 2 and 3, pl. 197, fig. 1, and pl. 198, fig. 2, 1903.

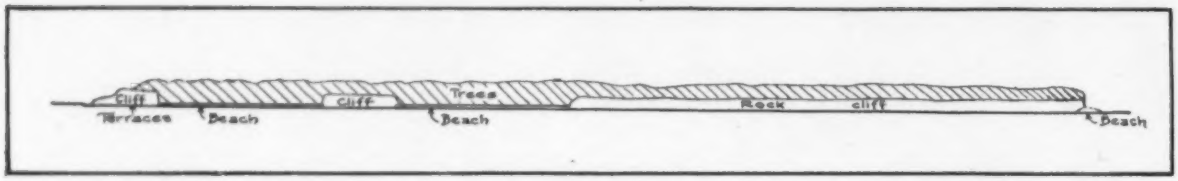


FIG. 19. The elevated atoll of Feys seen from the east. (Based on a sketch by the author.)

They are all steep-to islands and some of them are active volcanoes (map 17). The northern end of the line beyond Uracus is Farallon de Pajáras Reef, which is made up of lava (augite andesite).[23] The sections across the arc in the northern, middle, and southern portions are set forth in figure 20.

[23] Kaiser, Erich. Beiträge zur Petrographie und Geologie der deutschen Südseeinseln. *Jahrb. k. preuss. geol. Landesanst. u. Bergakad.* **24**: 114, 1903.

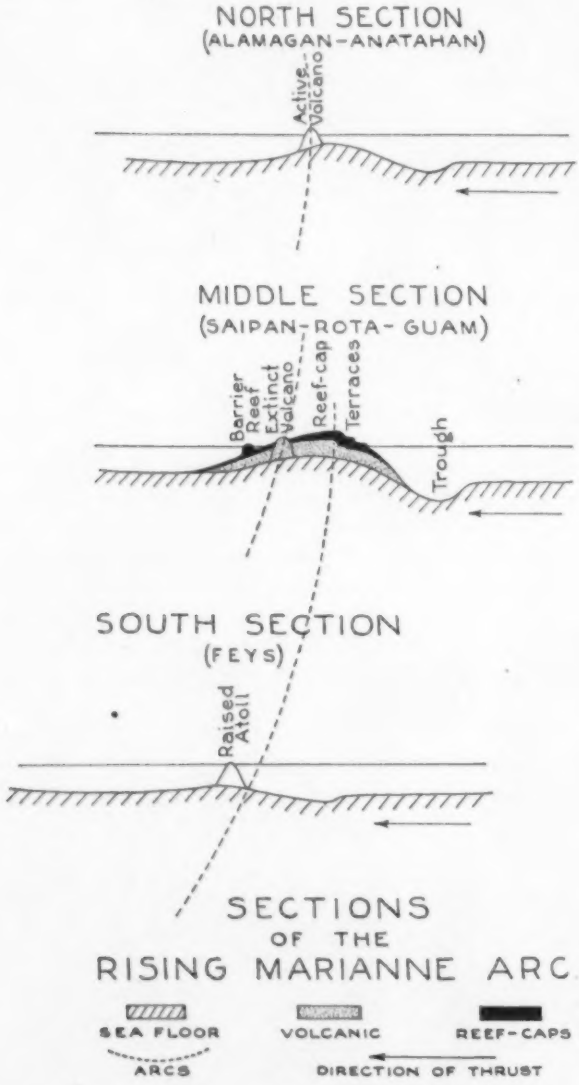


FIG. 20. Sections of the rising Marianne arc.

4. THE YAP ARC

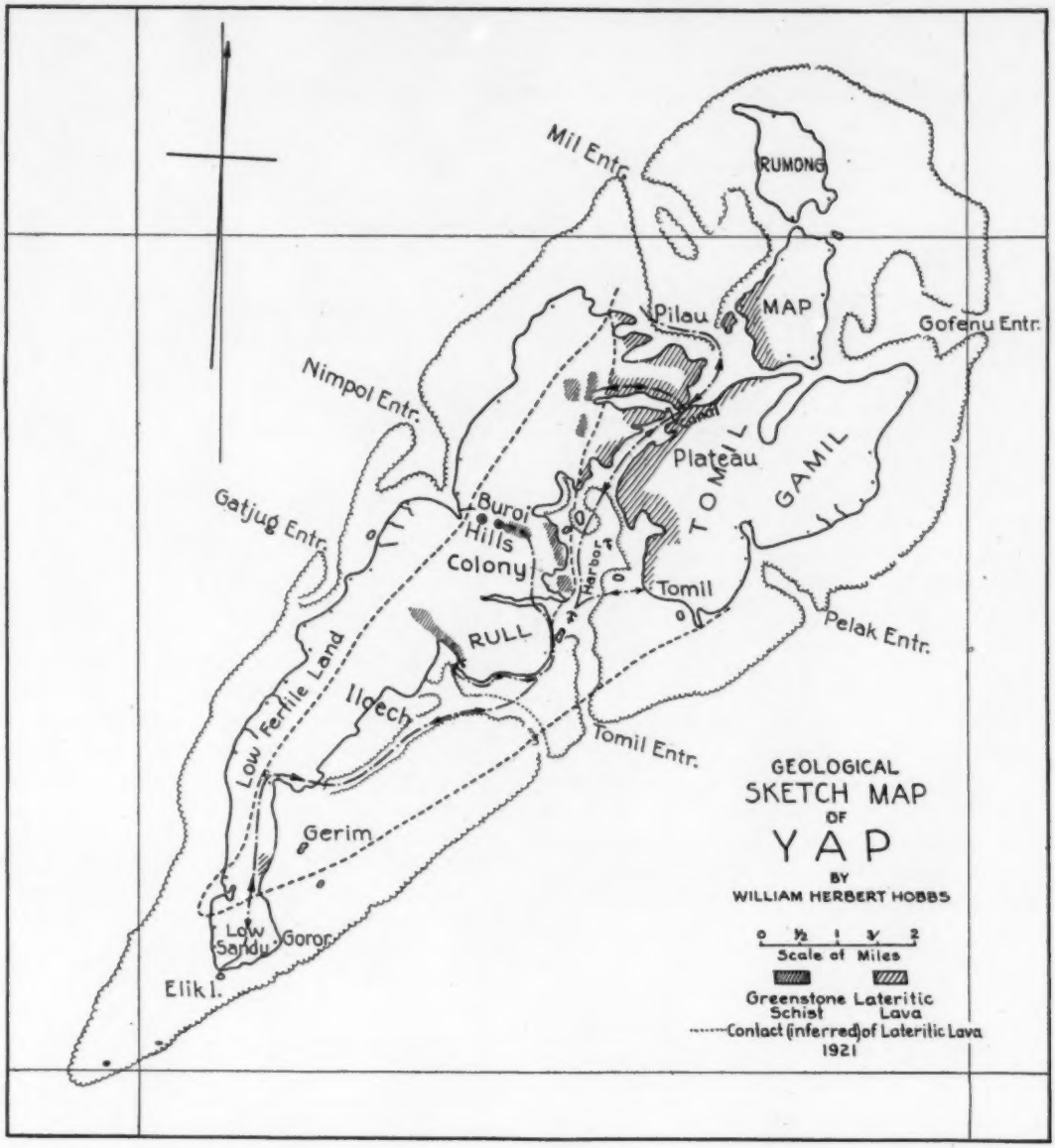
The Yap arc lies between the Marianne and the Pelew (Palau) arcs (map 17) and consists of what was clearly a single island. That it is in reality an arc despite its odd outline is revealed by the definite trough at its front.

At the present time Yap consists of four islands: Rull, Tomil-Gamil (separated from Rull by the Lagereng Canal), Map, and Rumong, all close together. Their separation is due to a recent subsidence of Yap, which drowned the young streams, formed shallow estuaries through the reef, and in places through the volcanic core of the original island as well (map 20).

Together the four islands make a triangular area which trends northeast-southwest for about 7 miles and has its shortest side at the north. The area has a general slope to the northwest (fig. 21*C* and *D*) and has a maximum elevation of about 500 feet.

Except for the estuaries, Yap is entirely surrounded by a reef intermediate between a fringing and a barrier reef, and this reef is generally more than half a mile and for considerable distances as much as 2 miles in width. This indicates that after the erection of the arc to above its present elevation there has been a long still-stand period, after which a quite moderate subsidence drowned the young river valleys and yielded the shallow estuaries through the reef.

Yap is composed wholly of igneous and apparently effusive rock, which at the surface presents two widely different types, both greatly decomposed. One is an older augite andesite, now a red lateritic decomposition product, which occupied the eastern half of the inland complex; and this is in surface contact along a north-south medial line with a basic greenstone schist, in places talcose or actinolitic, which is the decomposition product of a black basalt. Blocks large and small of the latter are strewn over the surface of the greenstone schist, and apparently this rock is in place in the canal. With much probability



MAP 20.

the contact of the intermediate and basic lava types along the medial line of Yap is one of superimposition of the basalt upon the andesite (map 20). According to both Kaiser[24] and Tsuboya,[25] there is to be found in the northwest island (Rumong) in a volcanic breccia blocks representing a considerable number of types of hypabyssal eruptive rocks related to the effusive types which so largely compose the island. These types are amphibole granite, amphibole syenite, gabbro, pyroxenite, and wehrlite.

[24] Kaiser, Erich. Einige Ergebnisse einer Reise nach den Karolinen und Mariannen. *Verh. XIII deutschen Geographentages zu Breslau im Jahre 1901*: 167–179, 1901.

——. Alte Gesteine von den Karolinen. *Zeitschr. deutsche geol. Gesell.* **54**: 62–63, 1902.

——. Beiträge zur Petrographie und Geologie der deutschen Südseeinseln. *Jahrb. k. preusz. geol. Landesanst. u. Bergakad.* **24**: 93–110, 1903.

[25] Tsuboya, K. Petrographical investigation of some volcanic rocks from the South Sea Islands. *Japanese Jour. Geol. and Geog., Trans. and Abstr.* 9 (3–4): 207–208, 1932.

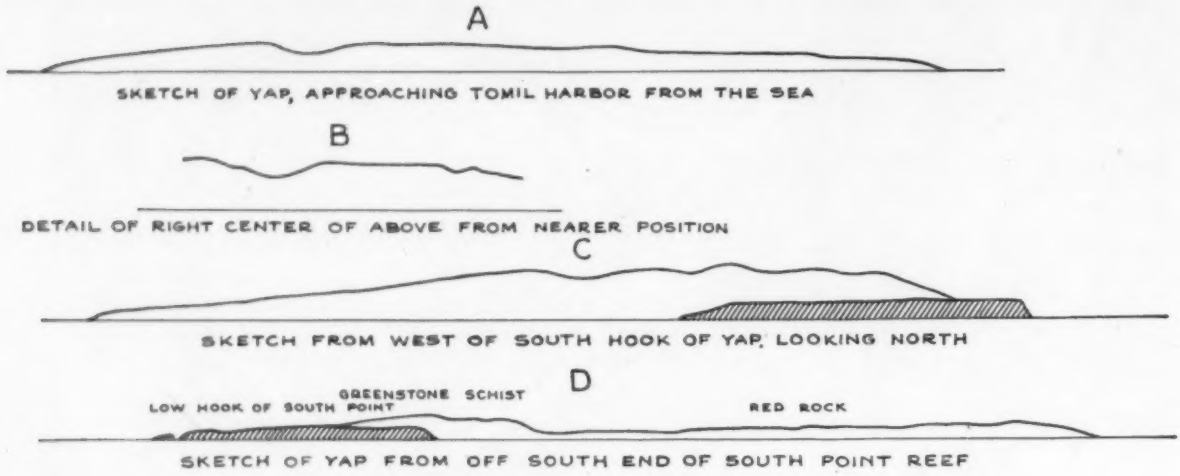

A
SKETCH OF YAP, APPROACHING TOMIL HARBOR FROM THE SEA
B
DETAIL OF RIGHT CENTER OF ABOVE FROM NEARER POSITION
C
SKETCH FROM WEST OF SOUTH HOOK OF YAP, LOOKING NORTH
D
GREENSTONE SCHIST
LOW HOOK OF SOUTH POINT
RED ROCK
SKETCH OF YAP FROM OFF SOUTH END OF SOUTH POINT REEF

FIG. 21. Profiles of the island of Yap. (From sketches by the author.)

The schistosity of the greenstone schist strikes northeast with the arc and dips northwest, hence parallel to the surface of the effusive rocks where uneroded. It is with much probability parallel also to an original fluxion structure.

As already pointed out, both the effusive rocks of the island are greatly decomposed. Erosion has been much more effective in removing the lateritic andesite, as is evident from the profile of the island as seen from South Point Reef (fig. 21*D*). It supplies confirmation of the long still-stand of the island, which is in striking contrast with the history of the Pelew arc next to the southwest.

5. THE PELEW (PALAU) ARC

The Pelew arc is in a more advanced stage of development than the Yap arc. It is about 100 miles in length, as against the 12 miles of Yap, and it subtends an arc of about 120°. The main central archipelago of the Pelews is a little more than half the length of the entire arc, since this is extended northward by the atolls of Kossol, Kriangle, and Aruangle (Vigaruangle) in that order (map 21, inset), and southward to the elevated atoll of Anguar or Nguar, which is reported to have an elevation of 65 feet.

The main island of Babelthuap[26] is in the north central or middle section of the arc, and is made up very largely of a hypersthene and augite andesite lava with its corresponding volcanic sediments, a red tuff largely altered to laterite and interbedded with agglomerate and volcanic breccia, the latter generally underlying.[27]

Babelthuap is about 27 miles long and 14 miles wide. Its crest is near the east coast, which is steep and terraced, as was clearly to be seen when we were cruising offshore. It is in this eastern zone that most of the solid lava is found exposed. From offshore this massive rock can be seen to be cut by dikes of a black and more basic rock. In contrast with the east coast, the west coast is gently sloping and also but little eroded (fig. 22*G*, *H*, and *I*). As on Yap, the considerable erosion is restricted to the high country, here toward the east coast. Residual spines of the lava are reported there to attain an elevation of about 2,000 feet.

Near the base of the gradual west slope of Babeldaob are cinder cones (fig. 22*G* and *H*), which quite likely belong in a line of volcanoes to the west of the first, which appeared near the east coast. In its higher levels Babeldaob is re-

[26] The name is also spelled Babeldaob and means "the island much elevated," a most appropriate designation which may indicate that the natives were observant.

[27] Oebbeke, K. Beiträge zur Petrographie der Philippinen und den Palau-Inseln. *Neues Jahrb. f. Mineral., Geol. u. Paläontol.*, Beil. Bd. **1**: 492–495, 1881.

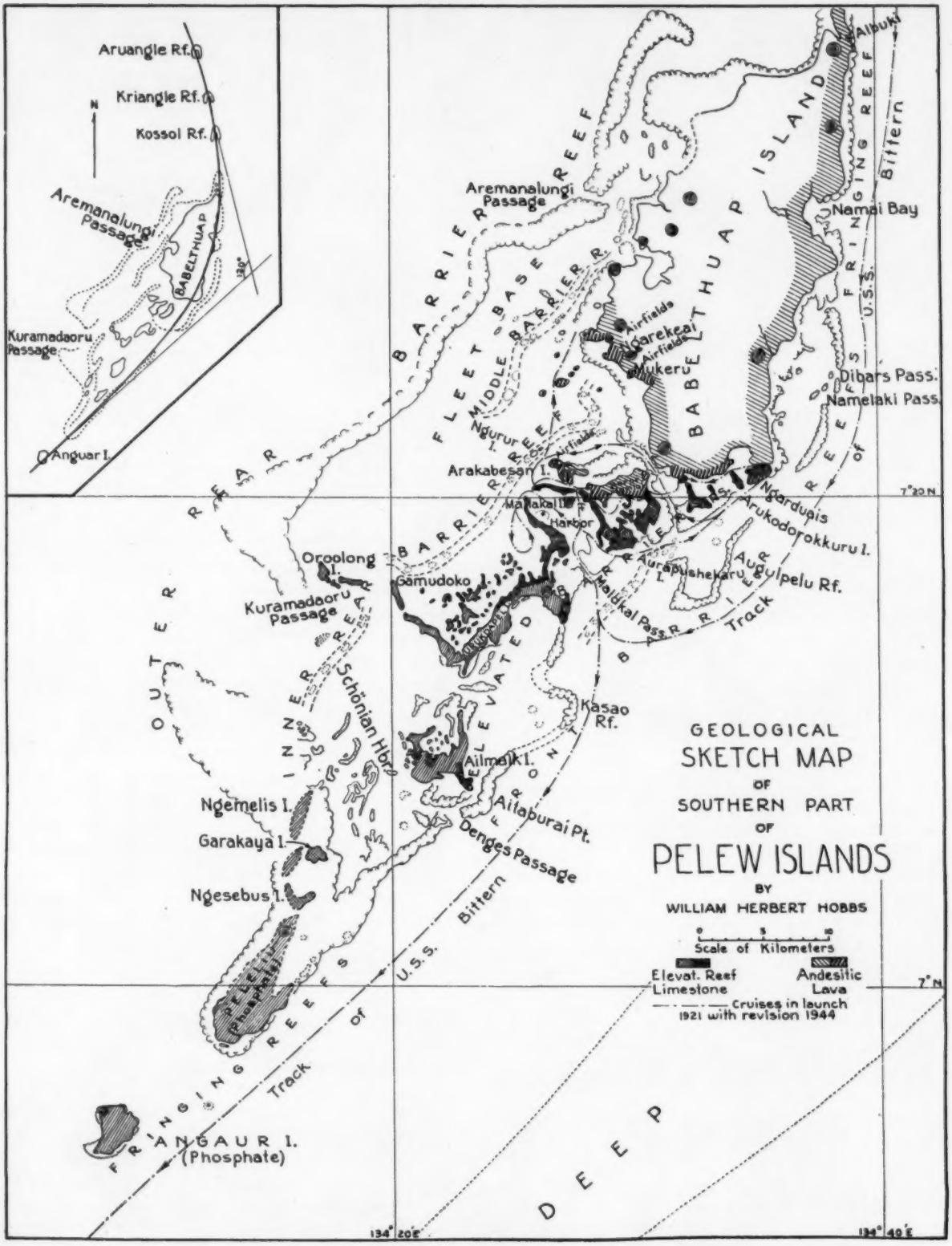
Kubary, J. Die Palau-Inseln in der Südsee. *Jour. Mus. Godeffroy* [Hamburg] **1**, Heft 4: 177–238, 1873.

Wichmann, A. Zur geologischen Kenntniss der Palau-Inseln. *Jour. Mus. Godeffroy* [Hamburg] **3**: 255–259, 1875.

Kaiser, Erich. Beiträge zur Petrographie und Geologie der deutschen Südseeinseln. *Jahrb. k. Preusz. geol. Landesanst. u. Bergakad.* **24**: 113–114, 1903.

Krämer, Augustin. Studienreise nach den Zentral—und Westkarolinen. *Mitteil. deutsch. Schutzgebieten*, **21**, Heft 3: 179–186, 1908. (Maps of Yap and Palau).

Tsuboya, K. Petrographical investigations of some volcanic rocks from the South Sea Islands. *Japanese Jour. Geol. and Geog., Trans. and Abstr.* **9**: 202–212, 3 pl., 1932.



MAP 21. Geological sketch map of southern part of Pelew Islands.
(Based on the Japanese hydrographic chart issued July 7, 1917.)[28]

[28] The U. S. hydrographic chart 1260 (1916) was also carried on the trip, but I was able to effect corrections in the position of the northern atolls, some of which are incorporated in hydrographic chart 5423 (1926).

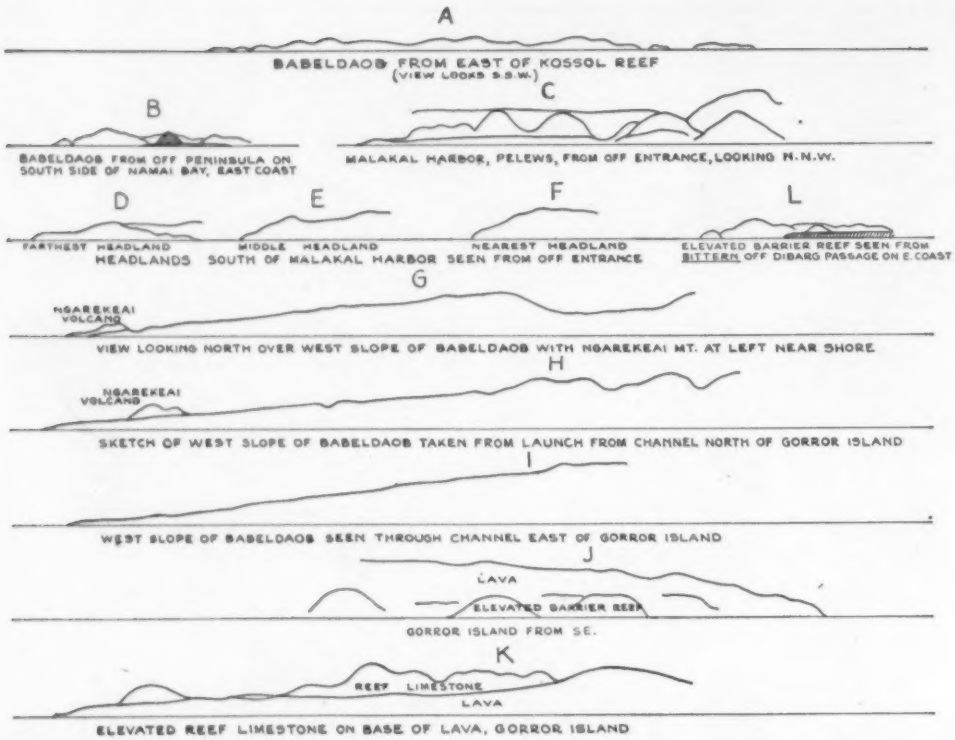

A
BABELDAOB FROM EAST OF KOSSOL REEF
(VIEW LOOKS S.S.W.)
B
BABELDAOB FROM OFF PENINSULA ON SOUTH SIDE OF NAMAI BAY, EAST COAST
C
MALAKAL HARBOR, PELEWS, FROM OFF ENTRANCE, LOOKING N.N.W.
D
E
F
L
FARTHEST HEADLAND
MIDDLE HEADLAND
NEAREST HEADLAND
HEADLANDS SOUTH OF MALAKAL HARBOR SEEN FROM OFF ENTRANCE
ELEVATED BARRIER REEF SEEN FROM BITTERN OFF DIBARG PASSAGE ON E. COAST
G
NGAREKEAI VOLCANO
VIEW LOOKING NORTH OVER WEST SLOPE OF BABELDAOB WITH NGAREKEAI MT. AT LEFT NEAR SHORE
H
NGAREKEAI VOLCANO
SKETCH OF WEST SLOPE OF BABELDAOB TAKEN FROM LAUNCH FROM CHANNEL NORTH OF GORROR ISLAND
I
WEST SLOPE OF BABELDAOB SEEN THROUGH CHANNEL EAST OF GORROR ISLAND
J
LAVA
ELEVATED BARRIER REEF
GORROR ISLAND FROM SE.
K
REEF LIMESTONE
LAVA
ELEVATED REEF LIMESTONE ON BASE OF LAVA, GORROR ISLAND

FIG. 22. Profiles of Babeldaob and other islands of the Pelew archipelago. (After sketches by the author.)

ported to be a barren trenched plateau which has been but little explored by geologists. However, there have been found overlying the tuff there, as well as near the west coast, ferruginous clays and a greenish-gray non-calcareous bedded marine sediment largely of organic origin.

The andesitic lava of Babeldaob is extended southward in the islands of Korror, Arakabesan, and Malakal, and as a basement to many of the other islands, which for most of their emergent portions are elevated barrier reefs. In these smaller islands also are exposures of an amygdaloidal melaphyre which may be related to the basic dikes seen cutting the eastern cliffs of northeast Babeldaob. Wichmann (*op. cit.*) reported also blocks of syenite granite, probably from volcanic breccia as was the case at Yap.

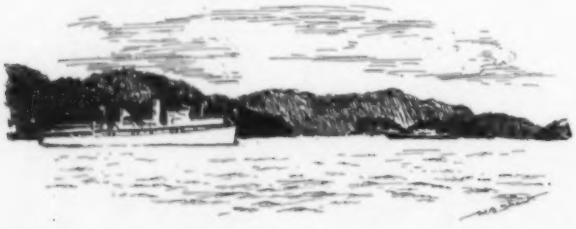


FIG. 23. The anchorage in the harbor of Malakal, showing high elevated barrier reefs at the back. In the foreground the U. S. mine sweeper *Bittern*, and at the right in the middle distance the *Chikuzen Maru*. (Pen sketch by Shaw after a photograph by the author.)

Encircling Babeldaob and the smaller lava islands are long chains of elevated barrier reefs, whose crests rise to elevations of 200 to 300 feet and which display the once submerged contours little changed, probably the most remarkable display of such structures to be found anywhere in the world (map 21 and figs. 22 (bottom), 23, 24*B* and *C*, 25, and 26). The channels between the reefs and the "sugar loaf" ranges are beautifully displayed. Though so little eroded, they reveal the stages in the interrupted process of elevation. Sea caverns cut in the sides of the reefs, cliffs, and even stacks are displayed in many of these elevated reefs.

My reading has revealed no description by a geologist of the extreme western reefs of the archipelago. We are therefore fortunate in having a rare account of these reefs by shipwrecked British seamen and their officers.[29] They found

[29] Keate, George. *An account of the Pelew Islands situated in the western part of the Pacific Ocean, composed from journals and communications of Captain Henry Wilson and*

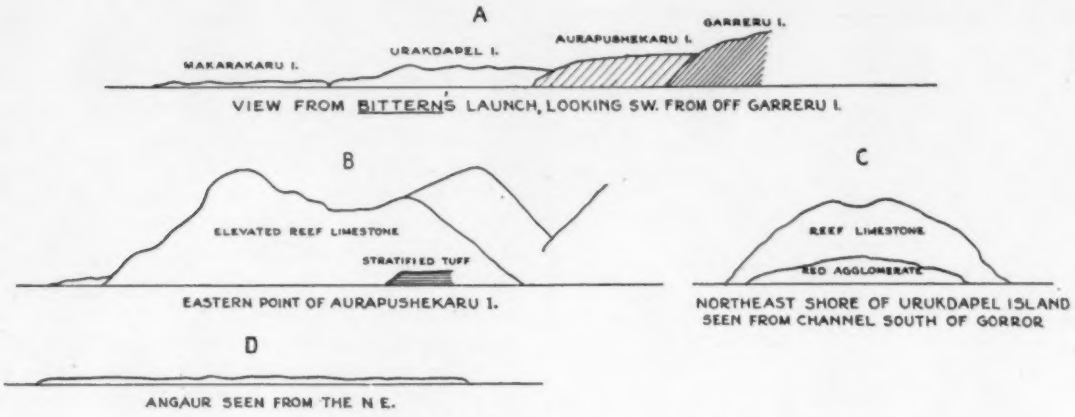


FIG. 24. Profiles of islands in the Pelew arc. (From sketches by the author.)

their way in the ship's boats to Oroolong Island, obtained supplies Crusoe fashion from the wreck, built a pinnace, and made their escape. They built an observatory on the island and from it made the careful sketch of the reefs to the eastward with the compass bearing of each "sugar loaf" (fig. 26).

The several elevated barriers (one at the front and at least two at the back), with the caves and cliffs upon them to mark the levels of ocean strands carved by the waves during their interrupted elevation, considered also with due regard to the fact that those in the middle front section are now being rapidly undermined by the waves during a subsidence—these facts reveal to us much of the late arc history.

The sequence of the vertical adjustments, in summary, seems to have been about as follows for the South Babelthuap - Korror - Malakal section of the arc:

1. Upward rising arc of andesitic lava with the usual terraces on front, and formation of barrier reef at the back. This later developed into

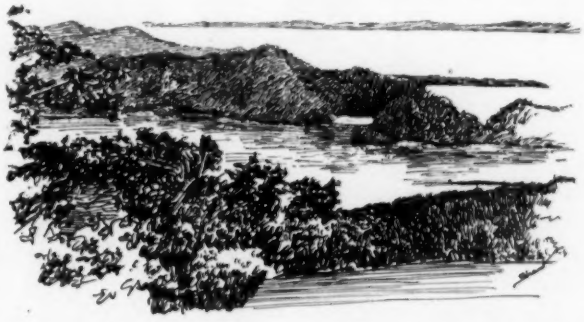


FIG. 25. View from the summit of Malakal island, looking northward over an elevated reef to Arakabesan island (andesite lava) in the middle distance, and beyond in the distance to the low west flank of Babeldaob (andesite tuff). (Pen drawing by Shaw after a photograph by the author.)

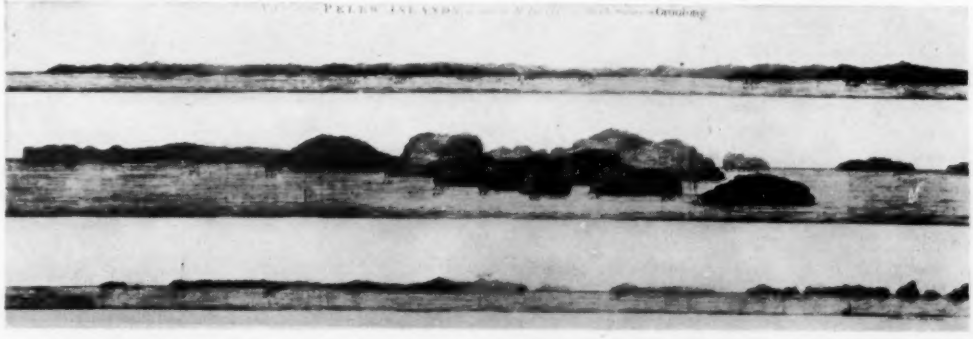
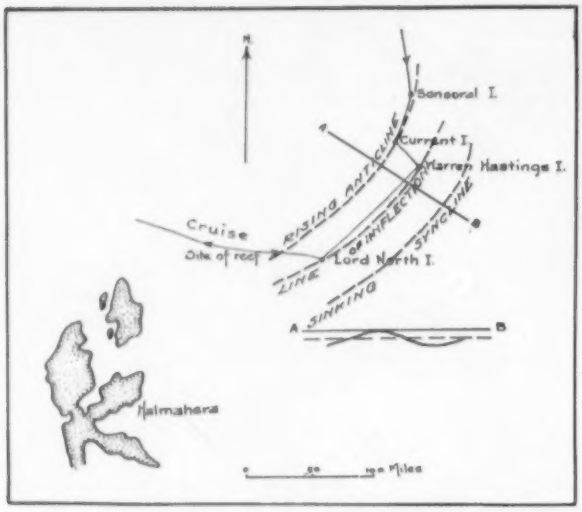


FIG. 26. Panoramic view of the elevated reefs of the inner barrier and of Gamudoko and Arukdapel Islands from the "observatory" on Oroolong Island, which is just inside the Kuramadaoru Passage through the Outer Rear Barrier Reef of the Pelew Islands. Date, 1783. (From Keate, 1788.)

some of his officers, who, in August, 1783, were there shipwrecked, in the Antelope, a packet belonging to the Honourable East India Company. London: 1–377, 1788. (See fig. opposite p. 288.)



MAP 22. The Sonsol group of atolls.

the longitudinal wrinkles (map 21, inset), with the atolls Kossol, Kriangle, the Aruangle at the north and the elevated Anguar at the south.

2. Later renewed lateral compression of the arc and development of barrier reef around the south of Babelthuap and its offshore lava islands, at the same time that a barrier reef developed off the front and a second one off the rear.

3. Renewed thrust from the front of the arc caused the islands with their encircling barriers to undergo an interrupted elevation.

4. The next succeeding thrusts came from the ends to produce further longitudinal wrinkling and a renewed subsidence in the cross synclines.

6. THE SONSOL ARC

To the southward of the Pelew arc and in the direction of the arcs of the Netherlands East Indies lies the inchoate arc of Sonsol. It consists of a few atolls and reefs. A still rather indefinite arc, it is made up (map 22), it would seem, of a front line consisting of Warren Hastings Island (also known as Merir and Pulo Marière) and Lord North Island (called also Neville and Tobi). The back line in the group consists of Sonsol Island (called also Sonsoral and St. André) and to the southwest of it Current Island (also known as Pulo Anna and Bur). Profiles of the islands made during a cruise near them are given in figure 27.

We approached the Sonsol group of islands in the *Bittern* from the north, intending to pick up Sonsoral on the port side. It was sighted on the starboard side, which may indicate that its position on the chart needs correction. The island is wooded and appears to be an elevated atoll. Our lookout in the crow's-nest said the island appeared hollow at the top, and I estimated its height to the tops of the palms to be about 100 feet. It is inhabited, and ten canoes of very wild-looking natives came out to the ship to trade.

Current Island is nearly circular and of about the same size as Sonsoral Island. It is a raised atoll, probably somewhat higher than that island. No natives appeared.

Warren Hastings Island is low, and the trees are wholly on a low reef of the east side, which is extended in a north-south direction. Lord North Island is similarly low, and the trees are on a north-south reef of the east side, but they extend also along the southeast side of the atoll as well. No natives were seen at Warren Hastings Island, but at Lord North Island five canoes of natives came out to barter. The Sonsol island group as a whole probably represents the earliest possible stage in the initiation of an arc.

7. THE RIUKIU (LOOCHOO) ARC

The arcs which have thus far been brought under consideration, from the Bonin arc at the north to the Sonsol arc at the south, comprise an outermost and latest series erected by late

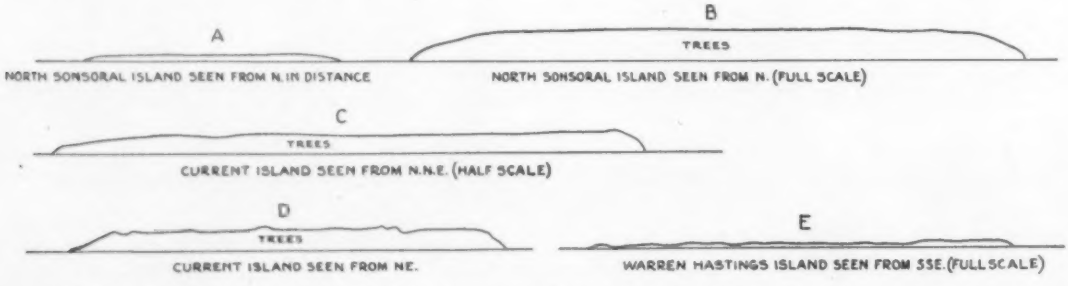
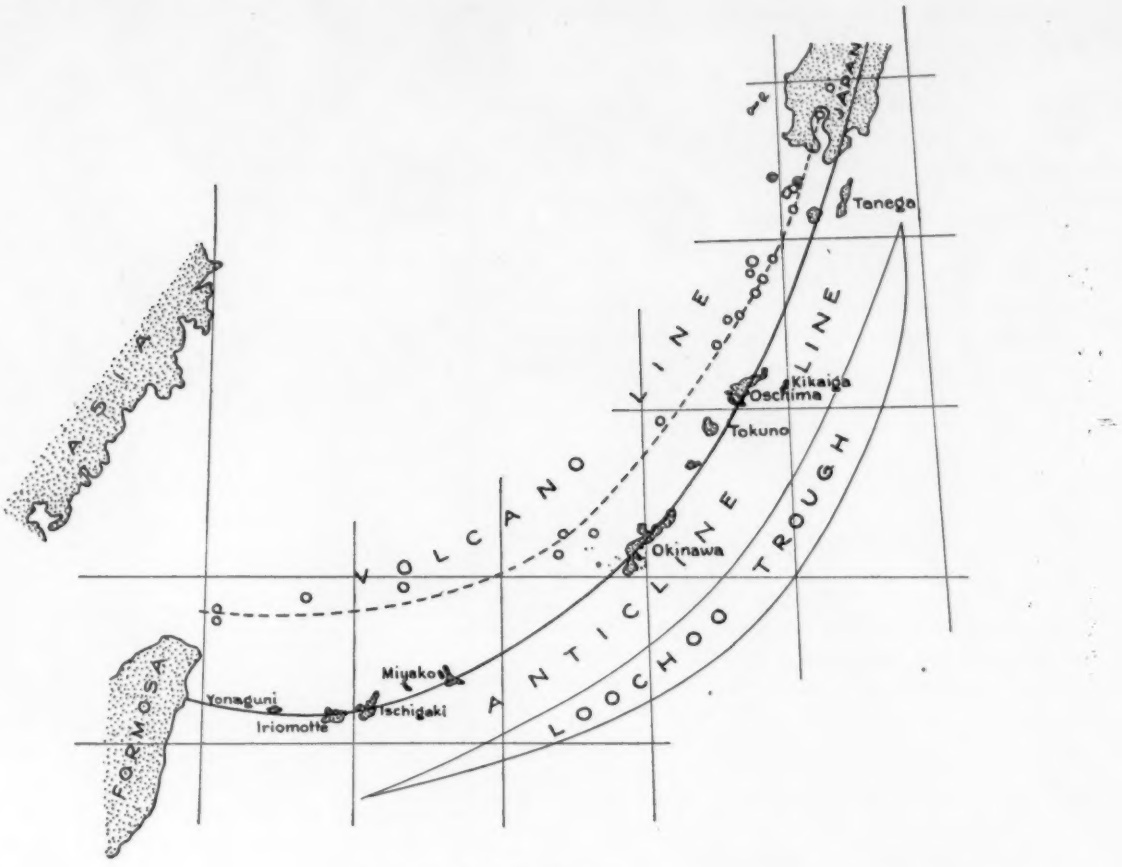


FIG. 27. Profiles of islands in the Sonsol group. (From photographs and sketches by the author.)



MAP 23. The Riukiu arc.

thrusts from the Pacific floor. An earlier series lies farther back toward the continent of Asia (maps 7 and 9), and they join the island of Kiusiu in southwestern Japan to New Guinea. In order they are the Riukiu (Loochoo) arc, the Philippine arcs, and the Sunda-Banda arc series. From northwestern New Guinea, where they are joined to the outermost series, they continue as a single series eastward through the Admiralty, Solomon, Santa Cruz, New Hebrides, New Calodonia, Fiji, and Tonga-Kermadec islands into New Zealand.

The Riukiu arc (map 23) is in a youthful stage similar to the Marianne arc. It joins Kiusiu to Formosa or Taiwan. Behind the deep Riukiu trough, a series of islands made up of Paleozoic and Tertiary sediments comprises the anticline crest, behind which in the normal position is a line of generally steep-to volcanoes which have erupted andesite lava. Of the sediments in the anticline, the later ones are at the front.[30] This front of the arc is strongly terraced with reef cappings, though neither barrier reefs nor atolls are at the back.

8. THE PHILIPPINE ARCS

Inasmuch as the outermost series of arcs off the Asiatic coast have so recently emerged from the sea, their surface features either betray little of the ravages of erosional agencies other than the notching by the waves, or they retain sufficient remnants of the anticline surface to permit in-

[30] Yoshiwara, S. Notes on the raised coral reefs in the islands of the Riukiu curve. *Jour. Coll. Sci. Imp. Univ. Tokyo* **16**: 1–14, 2 pl., 1901.

——. Geologic structure of the Riukiu (Loochoo) curve, and its relation to the northern part of Formosa. *Ibid.*: 1–67, 4 maps and tables of profiles.

Richtofen, Ferdinand von. Geomorphologische Studien aus Ostasien, III, Die morphologische Stellung von Formosa und den Riukiu-Inseln. *Sitzungsber. k. preusz. Acad. Wiss. Berlin* **11**: 944–975, 1 pl., 1902.

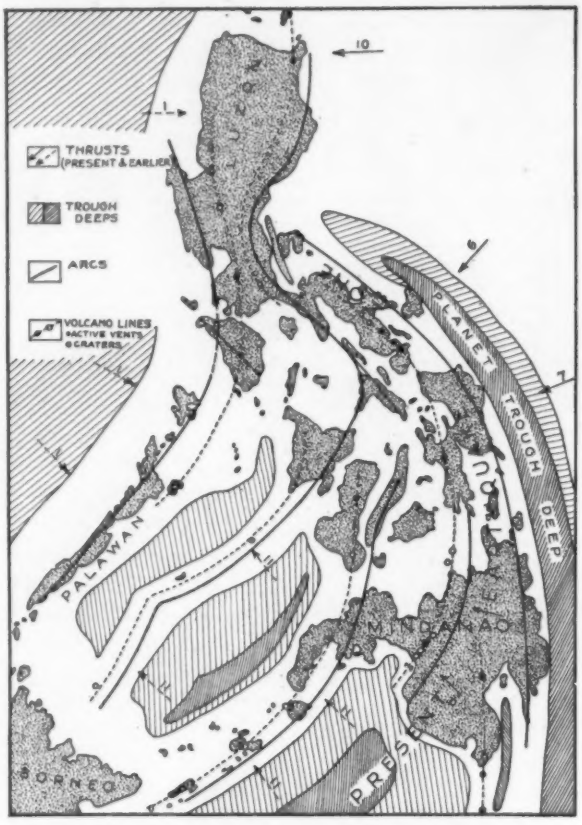
ferences to be safely drawn concerning their tectonic structure.

We turn now to an earlier and older series of arcs where erosional and transportational agencies have played a large part in obliterating original anticline surfaces, and where, in our attempts to decipher tectonic structures developed during mountain growth, we must rely more upon other factors, and especially on those learned from study of the younger and less modified arcs. Of these clues, those of greatest importance relate to the position of the lines of volcanoes both extinct and active, and to the position of troughs in the neighboring sea floor. The trends of the dominant ranges, with due regard to the strikes of the sedimentary rock formations, are of the greatest service in deriving the arcuate pattern, and this is very apt to be displayed in the trends of the relatively narrow islands.

Now that we feel assured in assuming that the thrusts which developed the arcs came from sinking sea floors at the front, and knowing, further, that the convexity, of the arc itself and the greatest convexity of the anticline profile both face seaward, we assign first importance to these facings. As regards the order of sequence in formation of arcs which are concentrically festooned, study of the zones of strong earthquakes should in all cases reveal the clue.

In now applying these criteria to the series of Philippine arcs, which with exception of northern Luzon belong in a single concentric arcuate group (map 24), we learn at the outset from study of the earthquake map[31] that the outermost arc next to the great Planet Trough Deep is the latest to have been formed (map 25). Probably the greatest earthquake intensity in the islands has been concentrated in the Agusan Valley north of Davao along the continuation northward of a narrow sea trough in the Gulf of Davao.[32] The Philippine rocks are all of Tertiary age or younger.

The arc which passes through the Bataan Peninsula in extreme western Luzon and is continued southwestward, passes through Mindoro and Palawan (the last two islands formerly were joined together). While it conforms in its arcuate convexity to the other arcs of the series, yet it is nevertheless abnormal in that the line of



MAP 24. The series of mountain arcs of the Philippines.

volcanoes, which includes the active volcano Taal southeast of Manila Bay, lies on the east (seaward side) of the anticline. It is the only case of its kind anywhere known to me. Luzon, in contrast to Palawan, is still very seismic (map 25), and the structure lines of Mindoro seem to run with the island or directly across the Palawan-Bataan structure.[33] There is here a suggestion that an early-formed arc ran southeastward with the Mindoro structure into Panay, the island to the southeast, which would conform to the early pattern, and that the late sinking of the great China Sea to the west, 1,000 miles across and with depths in excess of 5,000 meters, has superimposed a Bataan-Palawan arc upon the older one and so raised a line of volcanic vents on the back.

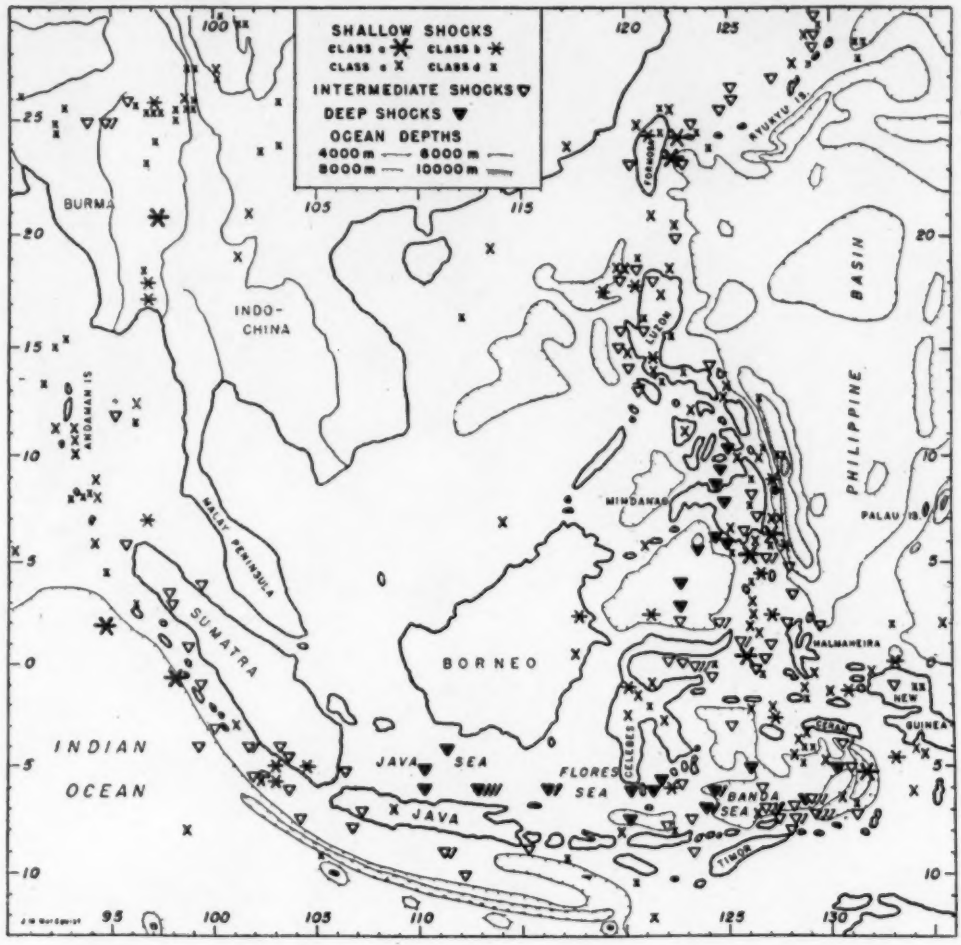
The other Philippine arcs, four in number, which have culminated in the one next to the

[31] Gutenberg, B., and C. F. Richter. Seismicity of the earth. *Geol. Soc. Amer., Spec. Paper* 34: 56, fig. 9.

[32] Smith, W. D. *Geology and mineral resources of the Philippine Islands*: 205–206. Manila, 1924.

[33] Smith, W. D. The essential features of the geology of the Philippine Islands. *Philippine Jour. Sci., A., Chem. and Geol. Sci. and Industries*, 5 (5): pl. II, 1910.

———. *Geology and mineral resources of the Philippine Islands*, pl. 37, 1924.



MAP 25. Earthquake map of the Philippine Islands and the Netherlands East Indies. (After Gutenberg and Richter.)

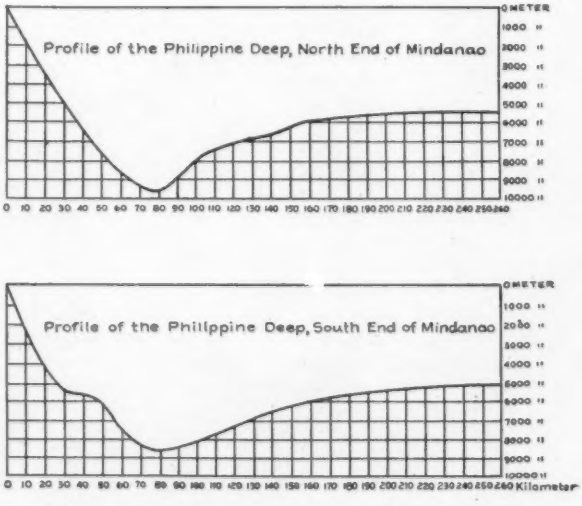

Profile of the Philippine Deep, North End of Mindanao
METER
Profile of the Philippine Deep, South End of Mindanao
Kilometer

FIG. 28. Profiles across the Planet Trough east of the Philippine Islands.

Planet Trough (fig. 28), are all quite normal, with trough to seaward and with the volcano line landward from the anticline.

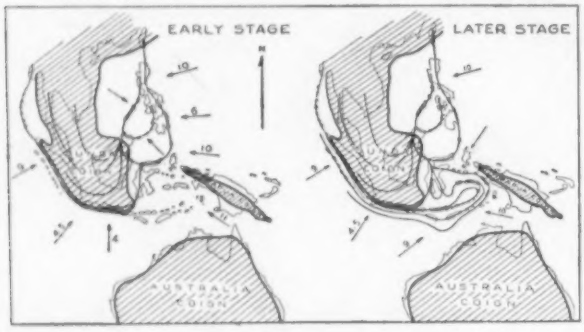
The Planet Trough has an extreme depth of 9,788 meters (31,811 feet) or over 6 miles. An interesting proof that this great depth has arisen through local subsidences is furnished by the sediments dredged by the ship *Planet* from the bottom. *Over* the deposit of minerals clearly derived from the augite andesite lavas was a 3½-inch calcareous deposit of brown clay containing well preserved shells of *Globigerina*. Since calcareous sediments are all dissolved before reaching a depth of 6,000–7,000 meters, the local subsidence has been in excess of 2½ miles.[34]

[34] Smolenski, J. Ueber die Entstehung der heutigen Tiefen des Philippinen Grabens. *Bull. Acad. Sci. Cracovie, Cl. Sci. Math. et Nat., Ser. A, Sci. Math.*: 586–601, Nov.–Déc., 1916.

All the early eruptions are of augite andesite, and there are later effusions of basalt.[35]

9. THE SUNDA-BANDA ARC SYSTEM

The Sunda-Banda arc system is unique, since it has been evolved, not as the result of a thrust from one ocean area toward a continent, but because of active thrusts from two oceans located on its opposite sides—a thrust from the Pacific basin lying to the northeast and another from the Indian Ocean off to the southwest (map 26).



MAP 26. The evolution of the Sunda-Banda arc system.

The Malay Peninsula, then as now, was extended southeastward by the Sunda Bank or platform which acted as a coign.[36]

In the Sunda part of the arc the thrusts appear to have come first from the Indian Ocean area, and these developed a double anticline on Sumatra and Java, the earlier one now being the most developed (map 27 and fig. 29). In addition to the two anticlines developed upon the larger islands there are two which are submarine ridges at the front. The inner one comes to the surface in the Mentawie chain of islands off central Sumatra, also in Soemba, Timor, and Ceram of the Banda region. Nias in the Mentawei chain is an anticline with the steep side on the south and with terraces on the north which are so recent that the color of the coral reef caps still shows.

The outer submarine ridge emerges in its eastern portion in Aroe and in the "bird beak"

[35] Iddings, J. P. The petrography of some igneous rocks of the Philippines. *Philippine Jour. Sci.*, Sec. A, **5**: 155, 1910.

[36] There is a voluminous literature relating to the structure of the Netherlands East Indies, beginning with the great works of R. D. M. Verbeek and coming down to works by the distinguished group of later official Dutch colonial geologists and mining engineers. Some of the more important of the later general works are the following.

On earthquake distribution:

Kemmerling, G. L. L. De aardbeving van Bali op 21 Januari, 1917. *Jaarb. Mijnwezen Nederl. Oost-Indië* **46**, *Verh.* **1**, 1917. Batavia, 1918.

Visser, S. W. On the distribution of the earthquakes in the Netherlands East Indian Archipelago, 1909–1919. *Verh. Kon. Magn. en Meteorol. Observatorium* [Batavia] **7**: 1–77, 3 maps, 1921.

——. Inland and submarine epicentra of Sumatra and Java earthquakes. *Ibid.* **9**, 1922.

Maps:

Abendanon, E. C. Geologische Schetskaart van Nederlandsch Oost Indië samengesteld in opdracht van het Koninklijk Nederlandsch Aardrijkskundig Genootschap. In six sheets. Scale 1 : 2,500,000. The Hague, 1914.

Atlas van Tropisch Nederland. (Issued by the Netherlands Geological Society in collaboration with the Topographical Service of the Netherlands East Indies.) The Hague, 1938.

De Bussy's Atlas Van Nederlandsch Oost Indië voor Kantoor, school en Hus. Amsterdam, 1919.

On structure of the archipelago:

Kotô, B. On the geological structure of the Malayan Archipelago. *Jour. Coll. Sci. Imp. Univ. Tokyo*, sec. **11**, pt. 2: 83–120, 1 pl., map, 1899.

——. The Rocky Mountain arcs in Eastern Asia. *Jour. Fac. Sci. Imp. Univ. Tokyo*, sec. II, **3**: 131–172, maps, 1931.

Suess, Eduard. *The face of the earth* (*The Oceanides*) **4**: 291–327. Oxford, 1909.

Volz, Wilhelm. *Der malaiische Archipel, sein Bau und sein Zusammenhang mit Asien. Sitzungsber. phys.-med. Soc. Erlangen* **44**: 177–204, map, 1912.

Molengraaff, G. A. F. Folded mountain chains, overthrust sheets and block-faulted mountains in the East Indian Archipelago. *Comptes Rendus Cong. Géolog.* Internat., 12th Sess., Canada, 1913: 689–702, 1914.

——. De geologie der zeeën van Nederlandsch Oost-Indië. In: De Zeeën van Nederlandsch Oost-Indië: 272–357, maps and plates. Leiden, 1921.

Brouwer, H. A. Geologische Oversichtskaart van den Nederlandsch Oost-Indischen Archipel. *Jaarb. mijnwezen* Nederl. Oost-Indië **44**, *Verh.* **2**: 272–354, 7 maps. Batavia, 1917.

——. The geology of the Netherlands East Indies, *Univ. Michigan Studies, Sci. Ser.* **3**: xii + 160, 18 pl., 17 fig., 1925.

——. The major tectonic features of the Dutch East Indies. *Jour. Washington Acad. Sci.* **12** (7): 172–185, 1922.

Gogarten, E. Die Vulkane der nördlichen Molukken. *Zeitschr. f. Vulkanologie, Ergänzungsb.* **2**: 1–298, 1 map, 1918.

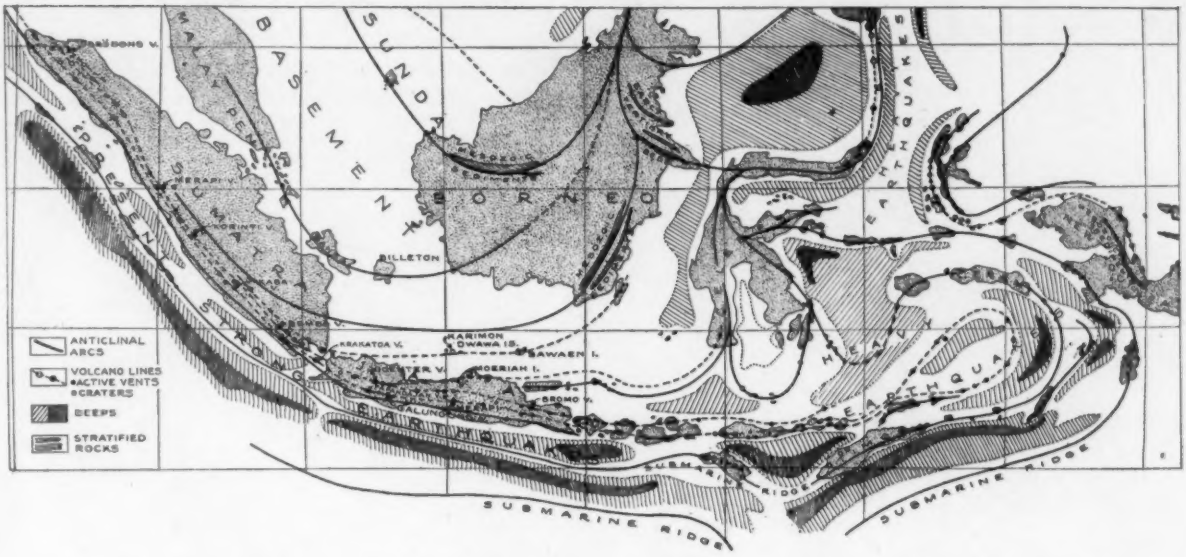
Hobbs, W. H. The unstable middle section of the island arcs. *Gedenkboek Verbeek Verh.* (*Supra cit.*, footnote 16).

Rutten, L. M. R. *Voordrachten over de Geologie van Nederlandsch Oost-Indië:* 1–839, illus. Groningen and The Hague, 1927.

Vening Meinesz, F. A. *Relevé gravimétrique maritime de l'archipel indien. Relation entre l'intensité de la pesanteur et l'activité tectonique de l'écorce terrestre.* (Publication de la Commission géodésique néerlandaise.) Pp. 1–6, map. Delft, 1931.

——. Modern deep-sea research in the East Indian Archipelago. *Geog. Jour.* **58**: 95–121, map, 1921.

Umbgrove, J. H. F. Over het Ontstaan van den Indischen Archipel. *Tijdschr. Kon. Nederl. Aardrijkskundig Genootschap*, ser. 2, **52**: 17–24, map, 1935.



MAP 27. The principal arcs of the Sunda-Banda system. (Based on *Atlas van Tropisch Nederland*, 1938.)

FIG. 29. Geological section across Java. (After Verbeek and Fennema.)

of New Guinea; thence it turns westward through Misool, Obi-eil, and Soela into the middle eastern prong of Celebes.

The lines of active volcanoes follow closely parallel to the latest formed anticlines—the submarine ridges which have today the heavy earthquakes. The gap in the line of active vents north of Timor is apparent only since they are covered by reef caps.[37] Farther back are lines of extinct vents which correspond to the older anticlines on the larger islands. The lavas from the series of young volcanoes near the southern coasts of the larger islands are exclusively andesites and rather generally augite andesites. The older volcanoes near their northern coasts, though in association with older andesitic eruptions, reveal that their later effusions were more varied and especially alkaline.[38]

The southern coasts of Java and Sumatra are marked by elevated strands in terraces, but these show downward as well as upward adjustments of the coast. On the Java south coast there is a strand of Miocene age at an elevation of nearly 1,000 meters and a strand of Upper Eocene age near present sea-level.[39] There are extensive barrier reefs off the north coasts of the smaller Sunda islands (Bali, Lombok, Soembawa, Flores, Soemba), but nowhere do they occur on the southern coasts.[40]

10. THE ARC SUCCESSION IN THE PHILIPPINE-SUNDA-BANDA ARC SYSTEM

As may be seen from maps 24 and 26, the Philippine and the Sunda-Banda arc systems are parts of a common complex, for the easternmost and latest anticline of the former, with its volcano line on the west, continues through the two

[37] Brouwer, H. A. On the non-existence of active volcanoes between Pantar and Dammer (East Indian Archipelago), in connection with the tectonic movements of the region. *Proc. Kon. Akad. Wetensch. Amsterdam* **21** (6–7): 795–802, 1917.

——. On reef caps. *Ibid.*, pp. 817–826, 1918.

——. Ueber Gebirgsbildung und Vulkanismus in den Molukken. *Geol. Rundschau* **8**: 197–209, 1917.

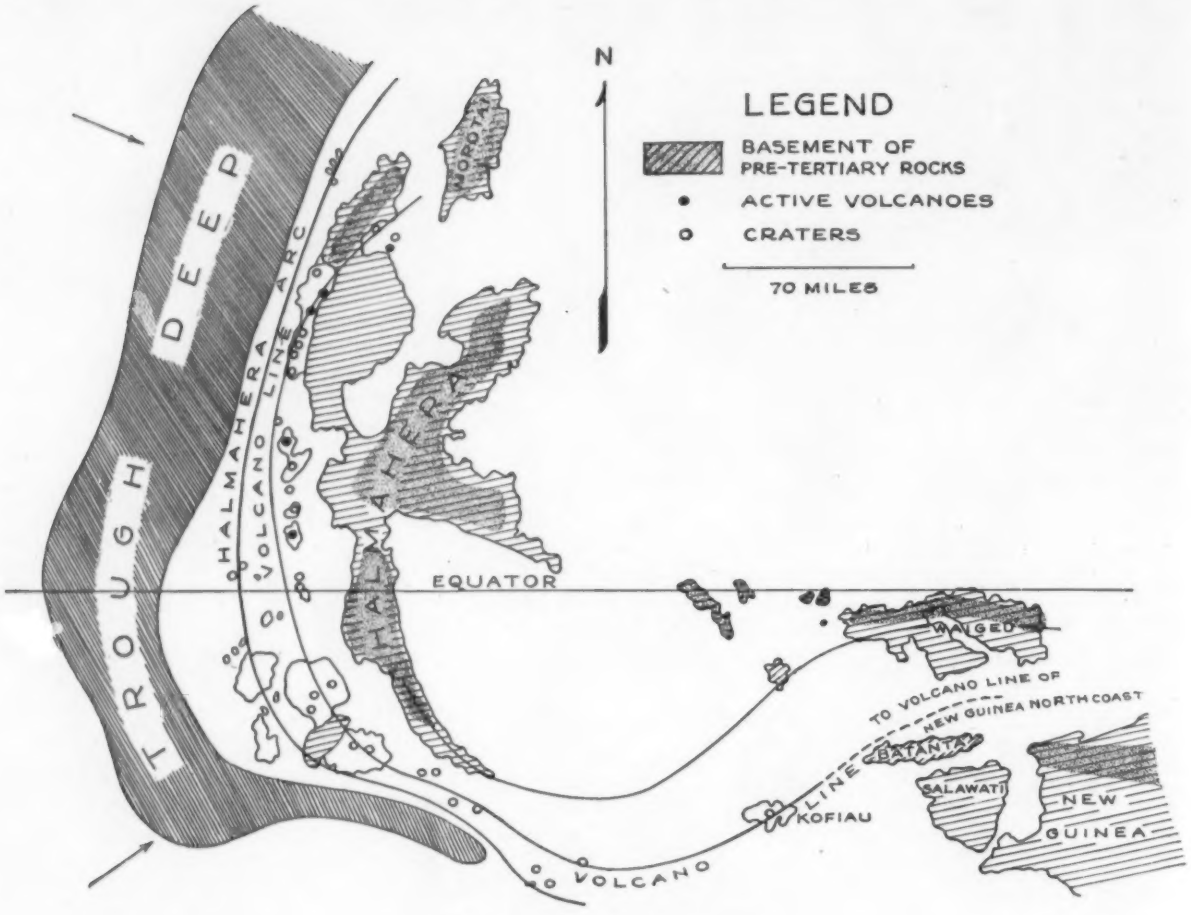
[38] Rinne, F. Beiträge zur Petrographie der Minahassa in Nord-Celebes. *Sitzungsber. k. preusz. Akad. Wiss. Berlin* **24**: 474–503, 1900.

Brouwer, H. A. Ueber leucitreiche bis leucitfreie Gesteine vom Gunung Beser. *Centralbl. f. Mineral., Geol. u. Paläont.* 1914 (1): 1–7, 1914.

[39] Martin, K. *Sammlungen des geologischen Reichsmuseums in Leiden*, N. F., Ser. I, **9**: 73, 1914.

——. Mesozoisches Land und Meer im indischen Archipel. *Neues Jahrb. f. Mineral., Geol. u. Paläontol.* **1**: 107–130, 1907.

[40] Niermeyer, J. F. Barriere-riffen en Atollen in de Oost-Indiese Archipel. *Tijdschr. kon. Nederlandsch Aardrijksk. Genootschap*, ser 2, **29**: 881, 1912.



MAP 28. The Halmahera arcs.

northeastern prongs of Celebes in the system of the latter. The arcs of Halmahera to the eastward are not represented in the Philippines, but form the connection with the outer series of arcs which started southward from the "elbow" of Japan and was continued through the Bonin, Iwo, Marianne, Yap, Pelew and Sonsol arcs to this junction with the Sunda-Banda system (map 28). They also form their junction[41] with the Bismarck Archipelago and its continuation eastward and then southward as far as New Zealand. The excessive compression of the Banda section of the Sunda-Banda arc system—it subtends an angle of no less than 212°—seems clearly to be due to thrusts coming mainly from the northeast and the southwest, with partial protection afforded at the front by the Papuan and Australian coigns (see map 26).

[41] Gogarten, E. Die Vulkanen der nördlichen Molukken. *Zeitschr. f. Vulkanol.*, Ergänzungsb. **2**: 1–298, 1 pl., 1918.

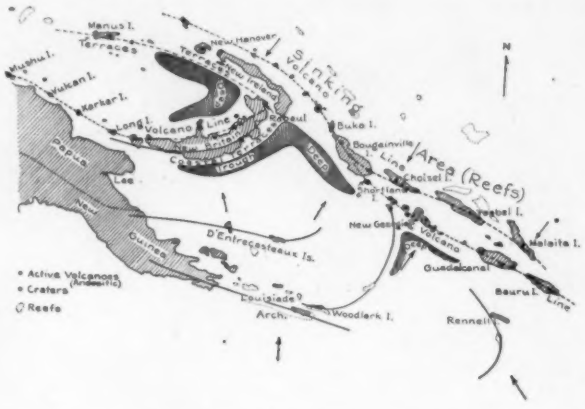
11. THE BISMARCK ARCHIPELAGO-SOLOMONS-NEW HEBRIDES ARCS

Comparable to the Banda arc in the measure of its compression is the Bismarck Archipelago[42] section of the arcs lying to the eastward, but the former was nearly symmetrical, whereas the Bismarck Archipelago has an entire lack of bilateral symmetry. The positions of anticlines, troughs, and volcano lines, and those of the terraces and reefs as well, clearly indicate that the arcs here face to the southeast, and that the major of the two opposing thrusts came from that direction (map 29).

Within these arcs there is the evidence that in late Pliocene to Recent times the islands have

[42] For the most detailed large-scale map of this archipelago, see R. H. R. Parkinson, *Dreisig Jahre in der Südsee, Land und Leute, Sitten und Gebraüche im Bismarckarchipel und auf den deutschen Salamoinseln.* Stuttgart: 1–876, 56 pl., 141 fig., 4 maps, 1907. (Double-page map of New Britain at p. 20 and of New Ireland at p. 248.)

risen from the deep sea a distance of at least a mile and a half, for marine deep-sea deposits containing shells of pteropods and *Globigerina* are found interbedded with the Recent volcanic tuffs and agglomerates. Today earthquakes bring about uplifts and subsidences equal to any



MAP 29. The Bismarck Archipelago - Solomons arcs.

in the world, and the coasts bob up or down in accelerated tempo.

When in the early 1880's Guppy visited the Solomons,[43] Treasury Island, which lies southwest of Choiseul, had been uplifted 5 feet during an earthquake, and a little earlier 7 to 8 feet. On Ugi Island, 400 miles away, there had been a drop of 15 feet. Earlier, however, the coast had risen to reveal a marine terrace at an elevation of 200 feet. On Treasury Island reef-capped terraces are at an elevation of 400 feet. Evan Stanley, after his second visit, this time with the Campbell Brown Expedition, stated that in the then recent earthquake the southern coast of New Britain had risen 20 feet, at the same time that the northern coast had sunk 2 feet.[44] In an earlier published report[45] he had written:

> New Britain is the axis of an old fold . . . which has suffered tilting from south to north. As the northern coast line subsided, volcanoes built up deposits of tuff, cinder and lava to an average height of about 3,000 feet. . . . Along almost the entire length of the southern coast line terraces of coral limestone and raised beaches may be seen up to altitudes of at least 1,500 feet.

The volcanoes erupted mainly augite andesites. Lavas and volcanic agglomerates were deposited and interbedded with marine sediments, Pliocene marls and grits, to indicate they erupted in the Pliocene and later. In New Ireland the sediments are folded "outward"—as in New Britain the facing is to the south.

In the Shortland Islands of the Solomons, pteropod ooze of deep-sea origin overlies the andesite lava, and this indicates that these islands must have been elevated from 4,500 to 6,300 feet, or probably about a mile. Bauru (San Christoval) Island at the extreme southeast end of the Solomons is 70 miles long and rises to an elevation of 4,100 feet. The coastal rim is a reef-capped wave-cut terrace 15 to 20 feet above sea-level, probably raised during a single earthquake, and such terraces extend upward to about the 500-foot level. On the south side the island rises precipitously, the front of the arc. The core of the island is of plutonic igneous rocks (diorite, gabbro, diabase, etc.), covered by volcanic extravasations.[46]

Bougainville Island at the opposite end of the Solomons is 110 miles long and 30 miles wide, and its volcanic peaks attain elevations of 7,000 to 10,000 feet.

Guadalcanal appears to have a core of plutonic rocks, like its neighboring island of Bouru, with volcanic cones reared above them. It is 80 miles long and 25 miles wide and has rivers flowing on flatter country on the northern side.

Florida Island, to the north of and near Guadalcanal, gives evidence by its excellent harbor that it has undergone a recent subsidence. It is largely made up of andesitic tuffs with reef cappings over foraminiferal limestones up to elevations of 900 feet.

Our knowledge of the New Hebrides is based largely on the explorations of Sir Douglas Mawson.[47] The history of these islands is one of almost continuous uplifts. According to Mawson, landslips in the soft rock of the mountains are almost continuous. Almost everywhere are elevated wave-cut terraces. On Santo they go

[43] Guppy, H. B. Observations on the recent calcareous formations of the Solomon group made during 1882–4. *Trans. Royal Soc. Edinburgh* **32**: 545–581, 2 pl., 1885.

——. Notes on the characters and mode of formation of the coral reefs of the Solomon Islands, etc. *Proc. Royal Soc. Edinburgh* **13**: 857–904, 1886.

——. *The Solomon Islands, their geology*, etc. London: 1–152, map and plates, 1887.

[44] Reported to the Second Pan-Pacific Science Congress convened in Australia in 1923.

[45] Stanley, Evan R. Notes on the structural relationships of the volcanic rocks, late Tertiary and Mesozoic deposits of New Guinea. *Rept. Australasian Assoc. Adv. Sci.*, **16**: 285, 1923.

[46] Marshall, Patrick. Oceania. *Handb. Region. Geol.* **7** (2), Heft 9: 17, fig. 6, 1911.

[47] *Proc. Linnean Soc. New South Wales*, 1905.

up to elevations of 1,000 feet, and in the Torres group they occur throughout, or up to 1,230 feet. Efate also is terraced to its crest at 700 feet. Between the New Hebrides and the Solomons lies the Santa Cruz island group with similar characteristics.

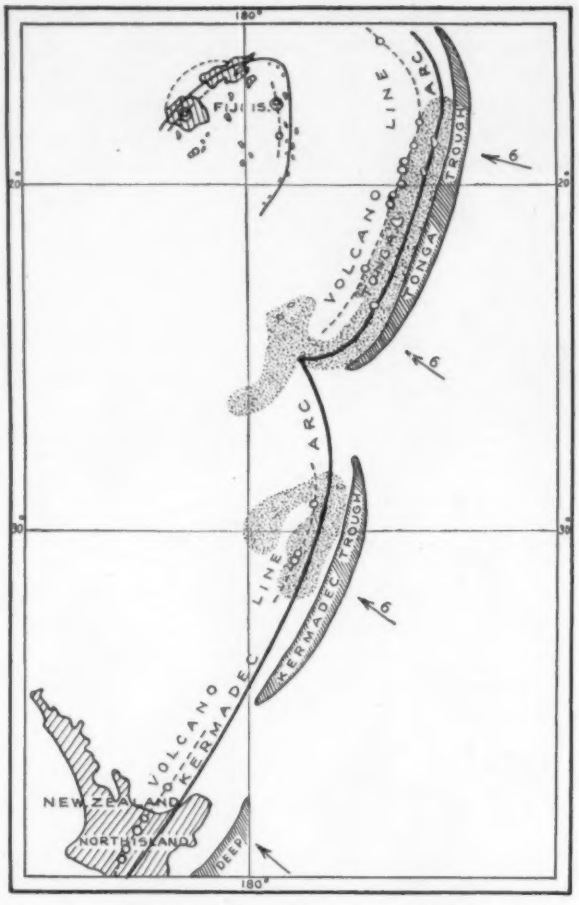
New Caledonia, which lies to the southward of the New Hebrides, is by contrast to them relatively stable. It represents the earlier anticline, while the New Hebrides constitute a near parallel one of more recent elevation, and now offer it protection from a major thrust which comes from the broad expanse of the Pacific lying to the northeast. Between lie the Loyalty Islands, and they, like New Caledonia, have a trough to the northeast as a facing. The Solomons at the west received their major thrust from the south and faced in that direction; but here farther to the east the major thrust was from the northeast with the facing of the arc in that direction. For the Solomons the thrust from the north may have been held off by the Carolines, Marshalls, and Gilberts, whereas for the New Hebrides no such shield existed.

12. THE TONGA-KERMADEC ARCS

The Tonga-Kermadec arcs to the northward of New Zealand face east-southeastward across a vast expanse of deep sea which is without obstruction as far as remote South America. In this respect these arcs resemble the Marianne arc, and like it they are in a youthful stage, are nearly rectilinear, have profound narrow troughs at the front and lines of steep-to volcanoes at the back (map 30).

Southward the Kermadec arc is continued into the easternmost arc of the North Island of New Zealand (map 31), and northward the Tonga arc probably makes connection with the New Hebrides section of the Bismarck Archipelago - Solomons - New Hebrides arcs (map 32), through a now submerged compressed arc like the Fiji arc and off its front.

Our knowledge of the Tonga geology is due largely to Lister,[48] and of the Kermadec arc, to Oliver and Speight.[49]



MAP 30. The Tonga-Kermadec arcs, with their relationships to New Zealand and to the Fiji arc.

The Tonga anticline takes its course in the normal way through a front line of islands which rise to an elevation of 500 feet and are composed of sediments, and these are faced by a staircase of terraces, reef-capped and with broad plateaus at the top. Between this frontal anticline and the line of steep-to volcanoes at the back, there is a middle line of islands made up of volcanic tuff laid down in the sea and lacking the frontal terraces of the front-line islands. Like the sediments of the Bismarck - New Hebrides anticlines, which were reared far out from a continental slope and had therefore to form their wrinkle on the floor of the deep sea, the sediments of the Tonga-Kermadec arcs are deep-sea sediments.

The Fiji, Tonga, and Kermadec arcs are all loci of heavy earthquakes which have accompanied the later adjustments in the anticlines. Angenheister,[50] who has mapped their distribu-

[48] Lister, J. J. Notes on the geology of the Tonga Islands. *Jour. Geol. Soc. London* **47**: 590–617, 1 pl., figs., 1891.

[49] Oliver, W. Reginald B. The geology of the Kermadec Islands, *Trans. New Zealand Inst.* **43**: 524–535, 4 pl., 3 fig., 1911.

Speight, R. Petrographical notes on the rocks of the Kermadec Islands. *Ibid.* **42**: 242–254, 1909.

[50] Angenheister, G. A study of Pacific earthquakes. *New Zealand Jour. Sci. and Tech.* **4** (5): 210–235, map, 1921.

tion, has shown that the strongest earthquakes have occurred on the undersea portion of the outer limbs of the Tonga and Kermadec anticlines. Another seismic zone of much importance is in the eastern wing of the Fiji arc. Samoa, near but located outside the belt of arcs and on the broad sinking floor of the Pacific, is relatively stable. The southwest wing (Vitilevu) and also that side of the Tonga arc (Tongatabu) have sunk.[51] However, it has been clearly shown[52] that the algebraic sum of the oscillations, upward and downward, in the Fiji group of islands since Pleistocene time has been an extensive uplift.

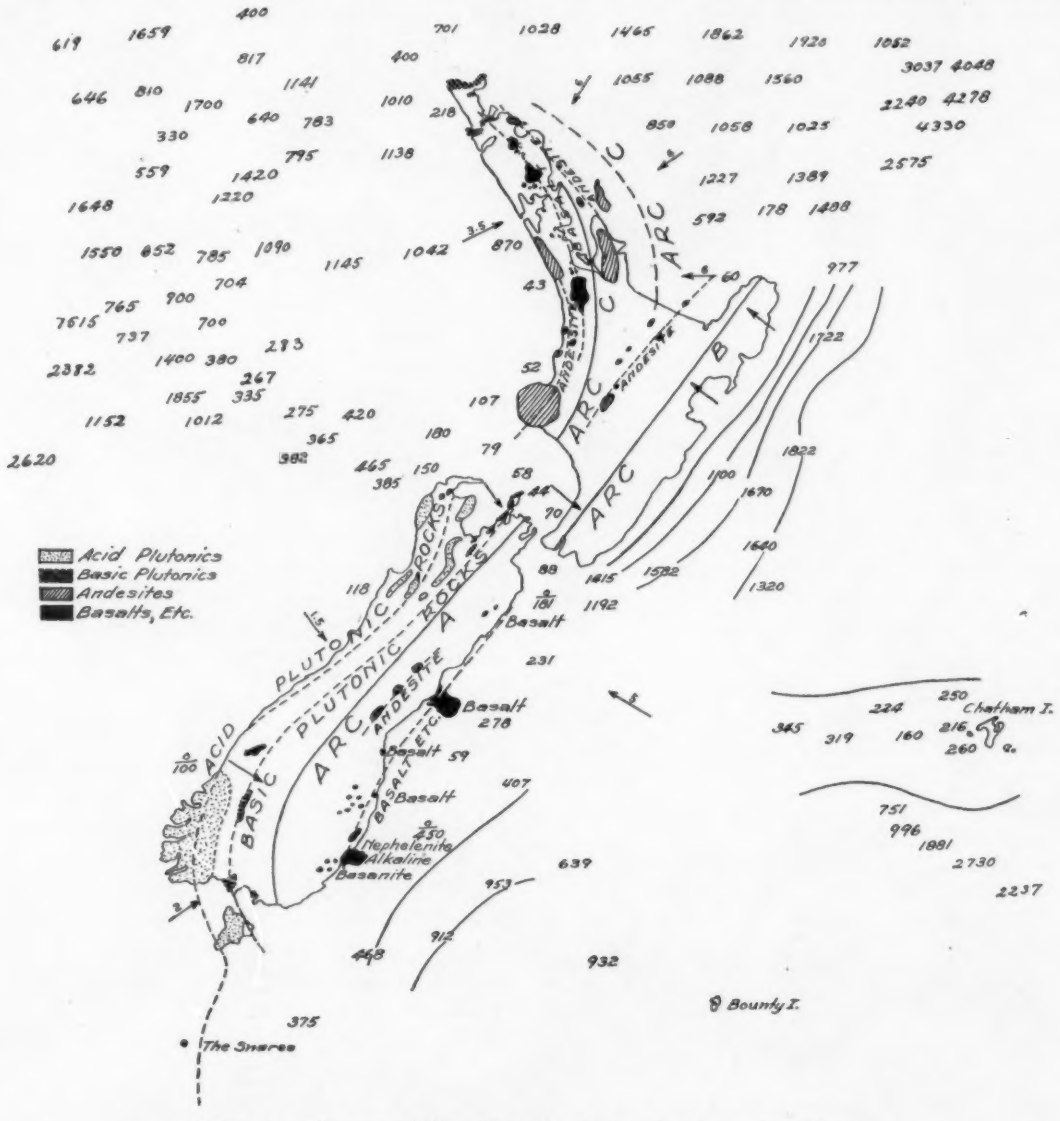
[51] Ladd, Harry S., and J. Edward Hoffmeister. Recent negative shift in the strand line in Fiji and Tonga. *Jour. Geol.* **35** (6): 442–556, 14 fig., 1927.

[52] Foye, W. G. Geological observations in Fiji. *Proc. Am. Acad. Arts and Sci.* **54**: 145, 1918.

13. THE NEW ZEALAND ARCS

New Zealand, a considerable land mass with mountains in excess of 12,000 feet elevation, and containing a core of sediments as old as the Cambrian, does not lend itself to a wholly satisfactory treatment of its mountain history by the methods of arcuate study alone.

The South Island, which exposes the core of ancient rocks within the deep fjords and valleys of the southwestern highland (Southern Alps), is



MAP 31. The arcs of New Zealand. Ocean depths are in fathoms.

somewhat set off from the North Island, which is connected up directly with the Kermadec arc only some 500 miles away. Here the dominant thrust which produced the arc came from the east (arc *B* of map 31), with a trough at the front and a line of andesitic volcanoes at the back.

The northwestern prong of the North Island by its convexity of curvature (dotted arc *C*) indicates that the thrust which reared it came from the northeast, and it faces there a great expanse of the Pacific, extending to the Americas. A line of lava extrusions from andesitic volcanoes is in a rear position. This anticline (dotted *C*), if ever pushed up above the sea, has since subsided; but behind it, following the course of a mountain backbone of the peninsula, is a parallel volcano line which has exuded basaltic lava of a later series. This is represented farther to the northwest by the corresponding granitic types of deeper-seated origin. A later and at the time dominant thrust came from the sea lying to the southwest. This might explain the andesitic effusions found along the southwest shore of the peninsula, and perhaps also some of the basaltic lavas along its medial line.[53]

The South Island is separated from the North Island by a great transverse fault in Cook Strait, along which the North Island has been shifted to the southeastward. In this older land mass to the south the structure reveals that the thrusts which raised the islands came in the northern section, where there was shielding from South Sea thrusts by the coign of Australia. Farther south no such shield was interposed, and heavy thrusts came in in a west-southwest direction from a vast expanse of sinking sea. This has produced a convex face of the arc *C* (map 31) in that direction.

The first anticline to rise produced the lofty range of the Southern Alps, now deeply dissected in the great fjords and deep valleys of the southwest. Of the volcano lines of this arc we can find only the roots—the granitic textured rocks of acid, basic, and ultrabasic stocks which, laid bare by the deep dissection, reveal the reservoirs only from which the earlier volcanic vents were supplied. In the secondary arc which was later erected at the rear (figs. 14 and 16), the front line of andesites and the rear one of basalts are both present, the latter following closely the line of the coast.

[53] For the geology and geography of New Zealand see the following works:

Park, James. *The geology of New Zealand*, etc. Christchurch: 1–488, 27 pl., 145 fig., col. geol. map, 1910.

Marshall, P. *The geography of New Zealand*. Rev. ed. Christchurch: 1–447, illus. (no date).

Cotton, C. A. *Geomorphology of New Zealand, pt. 1, systematic*. Wellington: 1–462, illus., 1922.

Marshall, P. The igneous rocks of New Zealand. *Gedenkboek Verbeek* (*supra cit.*, footnote 16): 357–367, 1925.

PART III, THE SINKING PACIFIC FLOOR

1. THE LAVA EFFUSIONS

With New Zealand, the long and in part double line of young mountain arcs beginning at Japan comes to an end at least so far as visible evidences are concerned. Across the South Seas beyond New Zealand lies that part of the Antarctic continent which, least known of the unknown, is yet thought to be like other parts of South Victoria Land, a block-faulted region.

To the eastward of the arcs which have now been passed in review lies the vast central area of the Pacific Ocean, in area nearly a third of the superficies of our planet. In sharpest contrast to the arcuate islands, those of the central Pacific area are without noteworthy alignment, though distinguished geologists, beginning with J. D. Dana, mixing arcuate islands with the others, have with rather indifferent success sought to arrange them in series of parallel straight lines regarded as fractures. All the islands are either volcanoes or atolls, now generally regarded as volcanoes, which have subsided as earlier reefs have risen over them to form the atolls.

When the lavas which these mid-Pacific volcanoes have erupted are subjected to petrographical and chemical study, it is found that they are in sharpest contrast to those from arcuate mountains. As we have seen, the arcuate mountain lavas of all earlier effusions are andesites, and nearly always pyroxene (augite) andesites. Sometimes a non-aluminous pyroxene (hypersthene) in part replaces the aluminous pyroxene (augite). It is the later extrusions only which in the case of arcuate mountains show variation from the rule that they are exclusively andesite—of intermediate chemical composition.

The lavas erupted from islands of the mid-Pacific are exclusively non-intermediate in composition; they are basic—basalts. Sometimes they are also alkaline as well, and, if so, the alkali constituents are either sodium or potassium, which are present largely in the minerals nephelene or leucite (see map 3).

Already convinced that the artificial distinction made by petrographers between "Atlantic"

and "Pacific" igneous rocks had been falsely based, in 1921, with the aid of Miss Ellen Stevenson, then instructor in geology at the University of Michigan, I assembled the chemical analyses of arcuate mountain lavas from the Pacific belt (776 separate approved analyses) and prepared the composite arcuate mountain lava.[54] This composite was found to be chemically closely in correspondence with the average mudstone or shale (see *antea*, p. 233, concerning a probable fusion of shale beneath a competent arch in a rising anticline).[55, 56]

The contrasts and parallels appear in the following table:

	Arcuate Mountain Lava				Fault Block Lava			
	1	2	3	4	1′	2′	3′	4′
SiO_2	58.29	59.09	56.69	57.76	49.52	47.62	47.92	42.74
Al_2O_3	16.97	15.35	16.47	15.22	15.89	15.20	13.59	11.13
Fe_2O_3 / FeO	6.94	6.88	2.67 / 4.87	6.74	11.50	4.50 / 8.85	2.80 / 9.67	3.96 / 9.66
MgO	2.83	3.49	5.07	2.76	5.04	4.90	6.04	11.63
CaO	6.06	5.08	9.06	3.07	7.27	8.42	9.20	13.04
Na_2O	3.55	3.84	2.08	1.25	4.03	3.20	2.98	2.64
K_2O	3.20	3.13	0.43	3.21	2.18	1.25	1.67	1.20
Others	2.16	3.14	2.23	9.99	4.57	5.69	5.96	4.35
	100.00	100.00	99.72	100.00	100.00	99.63	100.04	100.35

1. Arcuate mountain (fold) lava. Composite of 776 analyses.
2. Average magma. Composite of 5,602 analyses.
3. Andesite, Malakal Island, Pelews. Analyst, B. A. Soule.[56]
4. Average shale. Composite of 114 analyses.

1′. Fault block lava. Composite of 188 analyses.
2′. Basalt lava, Moen Island, Truk Archipelago. Analyst, B. A. Soule.[56]
3′. Basalt lava, Chogach Rock, Ponape. Analyst, B. A. Soule.[56]
4′. Basalt lava, From Lele Island, Kusaie. Analyst, R. K. McAlpine.[56]

A possible explanation of the basic character of the fault-block lavas is suggested by their origin in the adjustments in position of blocks within a mosaic at the time of earthquakes. Though in a sinking area where the average movement is a downwardly directed one, there are yet certain blocks which with regard to their immediate neighbors move upward, or at least downward by a smaller amount (fig. 30, lower profile). Moved relatively upward, the heavier upper strata of the block separate from the lower, only to be immediately caught in the jamming compressional thrust. This tends to hold the upper beds suspended as in a vice, lifting the load in part from the lower and inducing fusion. Later the load is applied to the fused mass, tending to lift it to the surface along a marginal fracture.

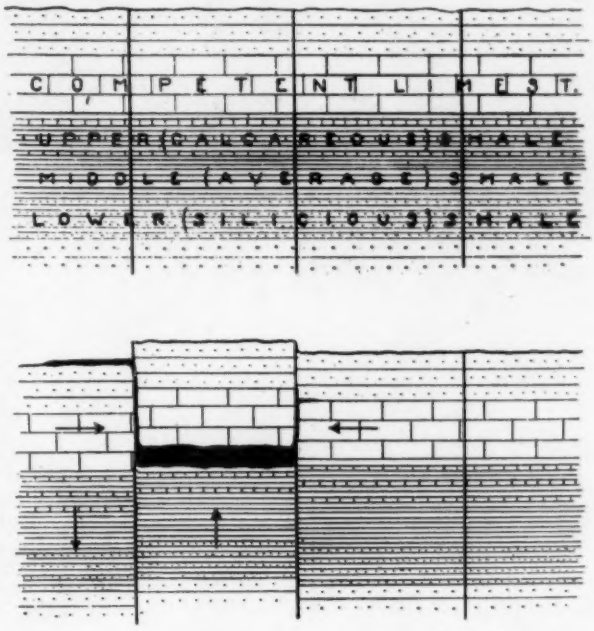


Fig. 30. Diagrams to illustrate a possible origin of fault-block lavas.

2. THE PARTITION LINE BETWEEN MOUNTAIN ARCS AND THE SINKING FLOOR

That a sharply defined boundary separates the area of arcuate mountains from that of the sinking Pacific floor off its front, was first recognized by Marshall,[57] and by the present author[58] it was extended to Japan (map 32).

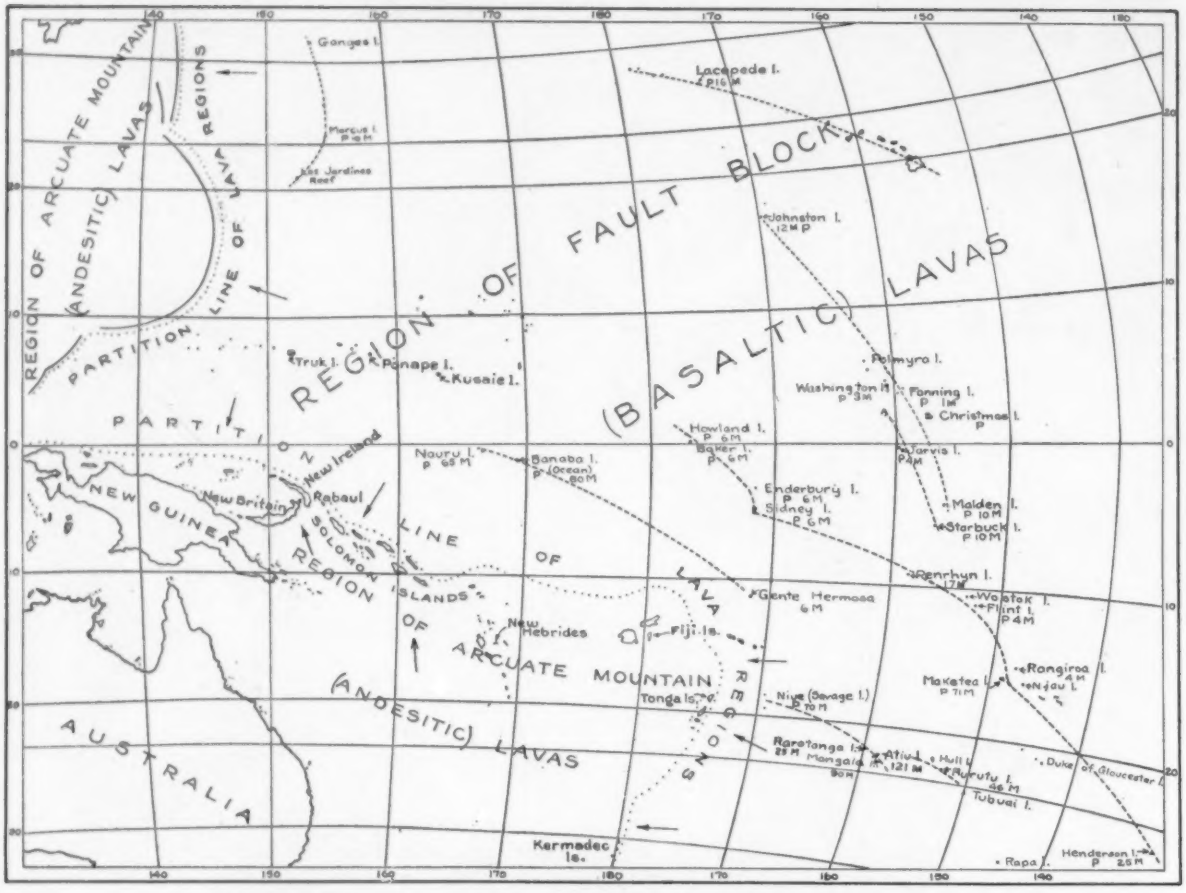
[54] Hobbs, W. H. Lava composition in relation to earth physiognomy. In: *Earth evolution and its facial expression*, ch. xiii, especially pp. 171–173, 1921.
——. Contrasted lava types of the Pacific region. *Second Pan-Pacific Sci. Cong., Australia*, 1923, 1: 838–842, 1926.

[55] Hobbs, W. H. *Earth evolution and its facial expression*, chap. iii. New York, 1921.

[56] Hobbs, W. H. and W. F. Hunt. Petrography of an area in the western Pacific. *Proc. Third Pan-Pacific Sci. Cong., Tokyo*, 1926: 715, 1927.

[57] Marshall, P. Ocean contours and earth movements in the south-west Pacific. *Rept. and Proc. Australasian Assoc. Adv. Sci., 12th Meeting, Brisbane, Sec. E:* 432–450, folding map, 1909.

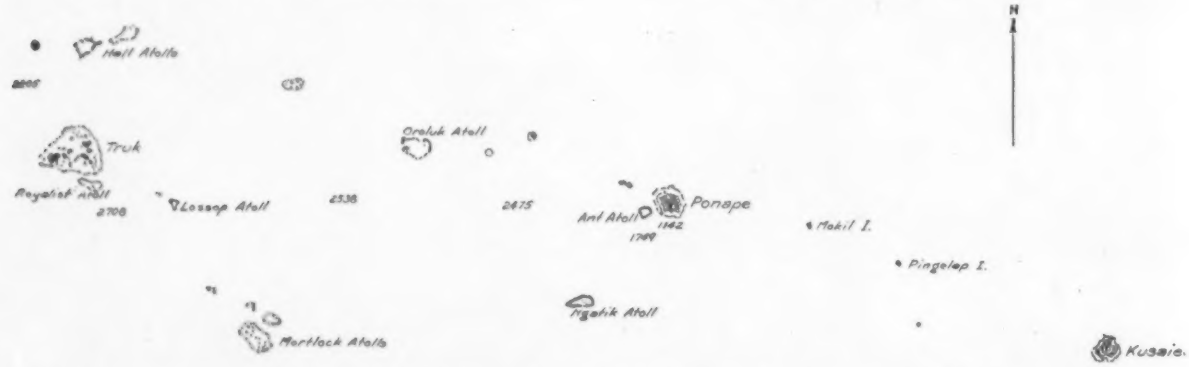
[58] Hobbs, W. H. Contrasted lava types of the Pacific region. *Proc. Second Pan-Pacific Sci. Cong., Australia*, 1923, 1: 842, 1926.



MAP 32. The regions of arcuate mountains with andesitic lavas contrasted with the region of fault-block adjustments with basaltic lavas.

3. BASALTIC VENTS IN THE SOUTHWESTERN PACIFIC FLOOR

Scattered throughout the broad expanse of the sinking Pacific floor are volcanic lava mountains which rise on steep slopes from generally profound depths. The slopes may in some cases be as high or higher than 30°. Of atolls, which are believed to be based on volcanoes, there are literally swarms, as, for example, in the Tuamotu or "Dangerous" Archipelago, and in those of the Gilberts and Marshalls. Much the most extensive of the lava islands are those of the Hawaian

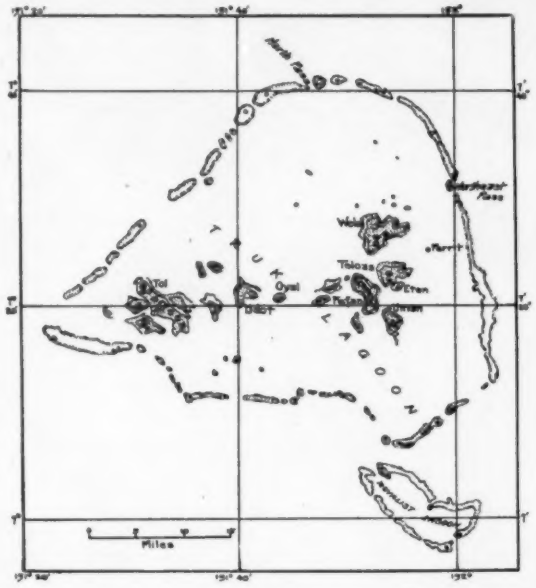


MAP 33. Map of the eastern Caroline Islands—the lava islands of Truk, Ponape, and Kusaie with associated atolls.

group, but the most important in the western section are those of the eastern Carolines—Truk, Ponape, and Kusaie (map 33).

4. TRUK

Of the basaltic volcanoes in the Caroline group, the largest is Truk, the great naval base of Japan. Among volcanic islands it is unique. It consists of an outer reef, the emergent portions of which rise at most 10 to 15 feet above the sea. The enclosed lagoon is roughly triangular, and 30 to 40 miles across. There are four entrances for vessels, two of which are indicated on map 34. The others are at the south and west.



MAP 34. Archipelago of Truk and the neighboring Royalist Atoll. The conical lava islands within the lagoon have fringing reefs. The lagoon has depths of generally more than 25 fathoms and less than 50 fathoms, and there are many reef islands.

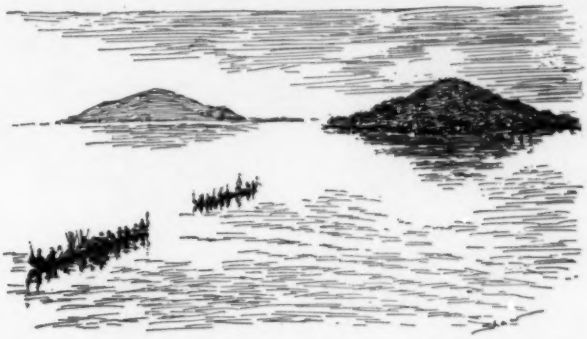


FIG. 31. Characteristic lava islands in the Truk lagoon. At the right is Eten, now truncated for an air field, and at the left Uman. (Pen sketch by Shaw from a photograph by the author made in 1921.)

The character of the ten islands of massive olivine basalt within the lagoon is indicated in figure 31. Other sectional panoramic profiles, made from the ship when east of the eastern passage through the reef, and from 5 miles north of the north entrance, are reproduced in figures 32 and 33, respectively.

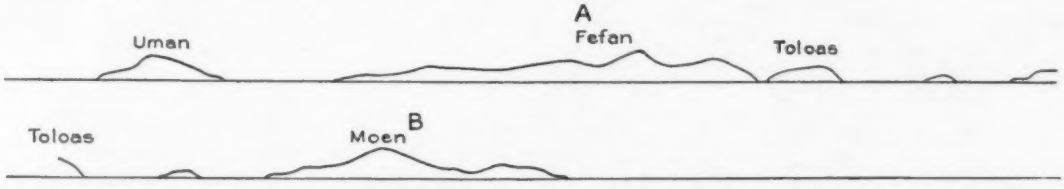


FIG. 32. Sketch of Truk, looking west from ship east of eastern passage. (*B* is continuation of *A* northward.)

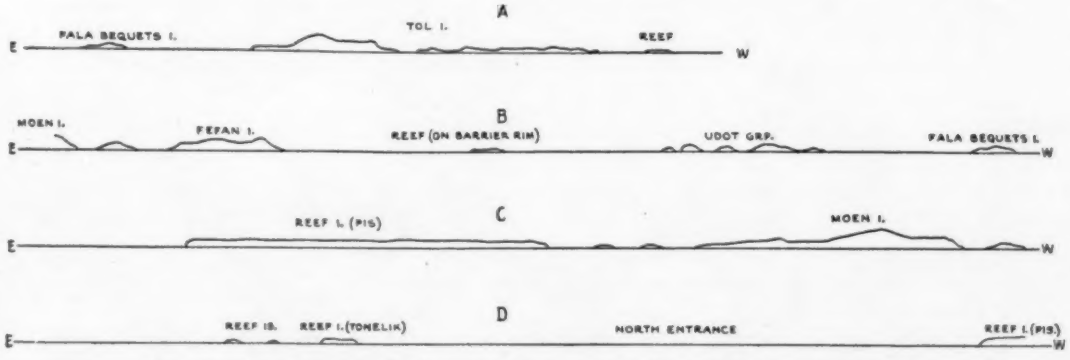


FIG. 33. Panorama of Truk from ship about 5 miles north of north entrance.

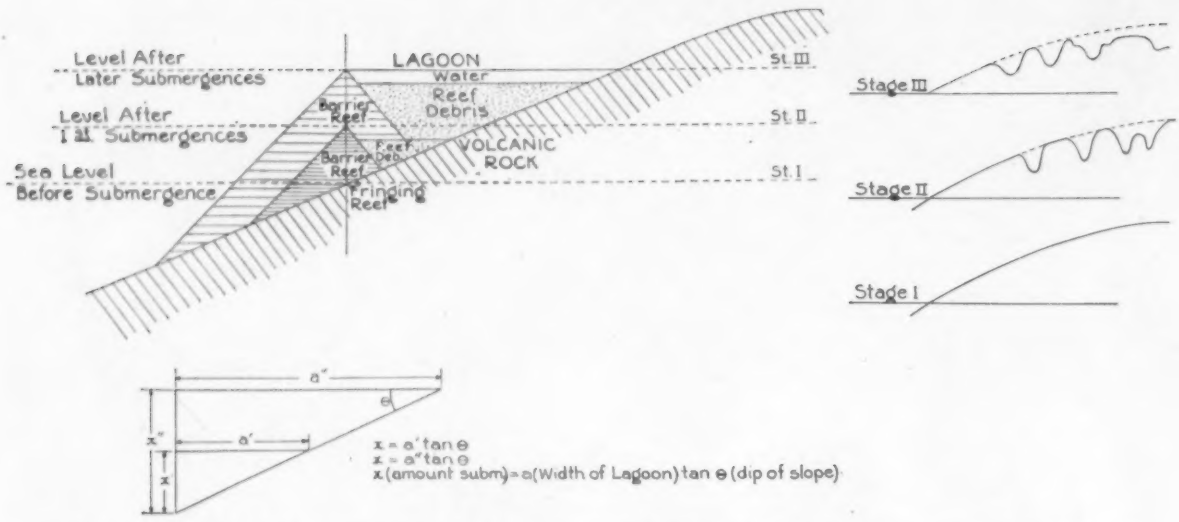


FIG. 34. Profiles of the encircling reef development and of the denudation of the lava dome of an earlier Truk Island, represented in three consecutive stages.

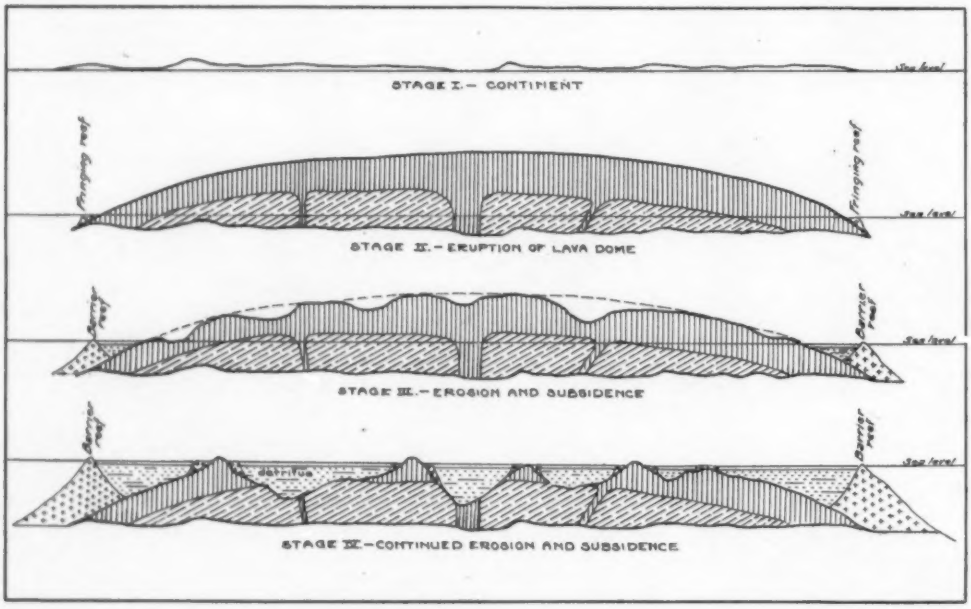
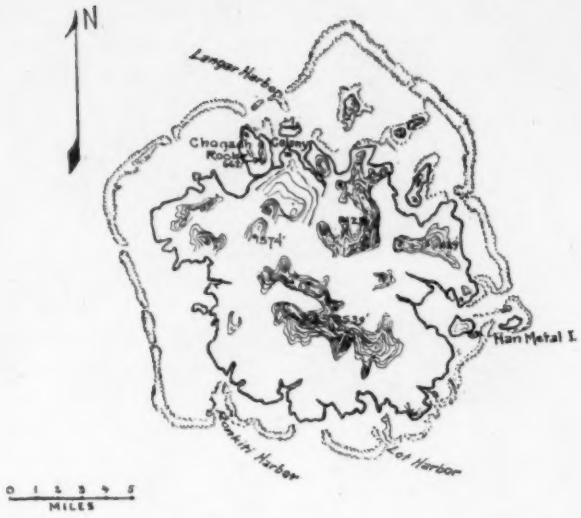

STAGE I.– CONTINENT
STAGE II.– ERUPTION OF LAVA DOME
STAGE III.– EROSION AND SUBSIDENCE
STAGE IV.–CONTINUED EROSION AND SUBSIDENCE

FIG. 35. Profiles to show stages through which an original Truk Island has passed to produce the present archipelago through reef formation, degradation of the lava dome, and subsidence.

The origin of the Truk Archipelago is explained through reef growth, denudation, and continued subsidence of volcanic extravasations in the form of a lava dome. Arbitrarily chosen stages I, II, and III in the formation of the encircling reef about the dome and of denudation are represented in figures 34 and 35. In figure 34 an original low continental area is represented as receiving large extravasations of basaltic lava in the form of a dome (stages I and II). Settlement and upgrowth of the encircling reef has produced the dissected island of stage III, and by further subsidence the archipelago of stage IV, which exists today.

5. PONAPE

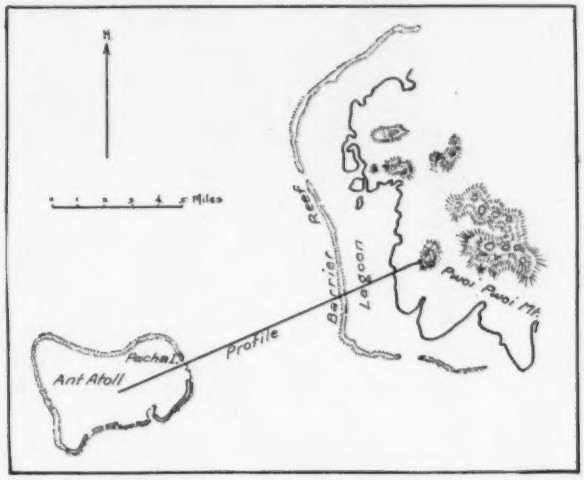
Ponape offers perhaps the world's most perfect example of an encircling or barrier reef. It sur-



MAP 35. The islands of Ponape in the eastern Carolines.

rounds this dissected lava dome, situated nearly on a direct line connecting Truk with Kusaie and somewhat east of the central position (map 35). Ponape has a maximum elevation of 2,579 feet and is made up of massive olivine basalt which grades into a nephelene basanite.[59] The principal entrance through the reef (and this admits large vessels) is at the north into Langar Harbor, the naval base under the dominating Chogach Rock, 937 feet high, which is cliffed on three sides, the north flank cliff actually overhanging (fig. 36*C*). Entrances through the reef which accommodate smaller vessels are off Ronkiti and Lot at the south of the island.

Compared with Truk, the degradation of the Ponape dome is small and the exudations which built it up must have been much later. This is confirmed by the lagoon width, which here averages only 2½ miles.



MAP 36. Western Ponape with Ant Atoll.

Off to the west-southwest of Ponape a distance of less than 10 miles is the perfect atoll of Ant (map 36). Because of the interesting relationship here displayed between a typical barrier reef and a typical atoll, soundings were undertaken and a section was made which extends from the summit of Pwoi Pwoi Mountain on Ponape to the center of the Ant lagoon (fig. 37).

[59] Hobbs and Hunt, *op. cit.*

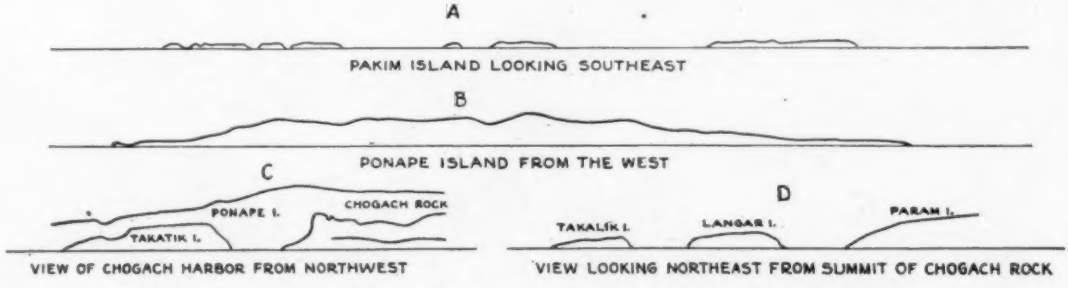


FIG. 36. Profiles of the island of Ponape. (From sketches by the author.)

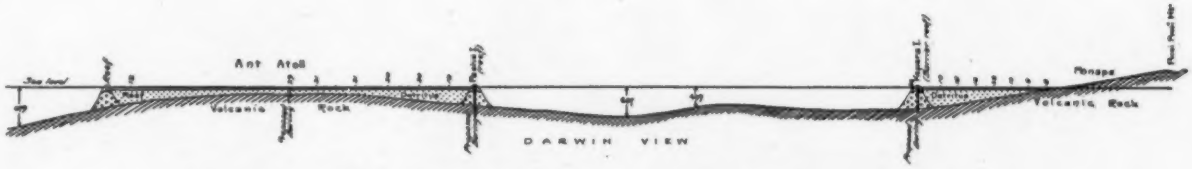


FIG. 37. Section from Pwoi Pwoi Mountain on Ponape to the center of Ant lagoon.

Also to display the striking contrast between the large degradation on Truk and the relatively small amount which Ponape has undergone, the profiles of figure 38 are given.

Though I did not ascend to the highest point of Ponape, I was convinced from conversations with Chief Nanpei of the Ronkiti tribe of natives, an educated native who speaks English, that the island is what its profiles proclaim it, a craterless lava dome. Its probable later history is indicated in figure 39.

6. KUSAIE

The island of Kusaie I did not visit, and I am therefore compelled to depend upon others for such knowledge as I have of it. The profiles in figure 40 are from photographs by various visitors to the island.

The lagoon of Kusaie is narrow, ½ mile to a mile in width, as against the 2½-mile-wide lagoon of Ponape and the very much wider one of Truk. The summit of Kusaie is relatively flat-topped and at an altitude of 2,100 feet. The island is, like Ponape, nearly circular in outline, here with an extension in an east-west line of 7 miles, and in a north-south line of 5 miles.

Captain U. Aoki, in command of the *Matsuyama Maru*, was good enough to collect a specimen of the lava from Lele Island near the harbor of Kusaie. This has been studied in thin section

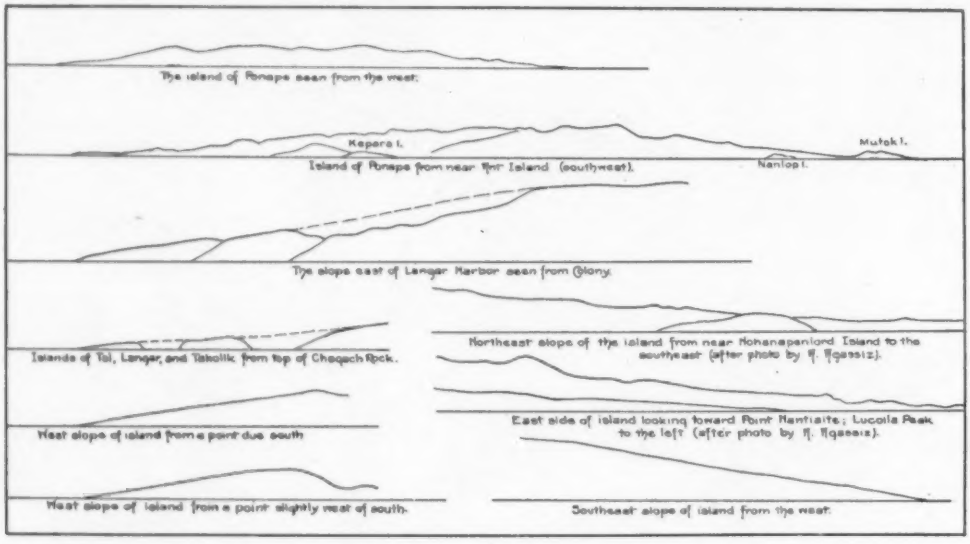

The island of Ponape seen from the west.
Kepara I.
Mutok I.
Island of Ponape from near Nut Island (southwest).
Nanlop I.
The slope east of Langar Harbor seen from Oiony.
Northeast slope of the island from near Mohnapanlond Island to the southeast (after photo by H. Hagesis).
Islands of Tol, Langar, and Takolik from top of Chogach Rock.
East side of island looking toward Point Nantiaite; Lucoila Peak to the left (after photo by H. Hagesis).
West slope of island from a point due south
West slope of island from a point slightly west of south.
Southeast slope of island from the west.

FIG. 38. Profiles of Ponape. (From sketches by the author.)

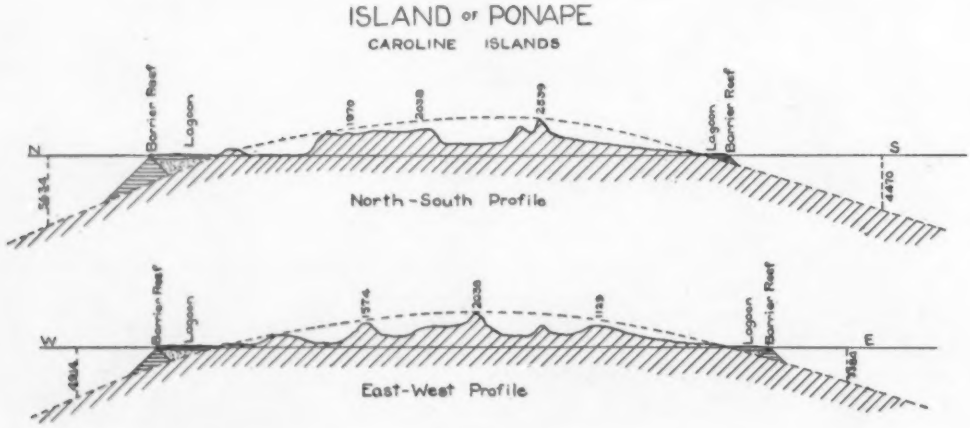


FIG. 39. Sections across Ponape.

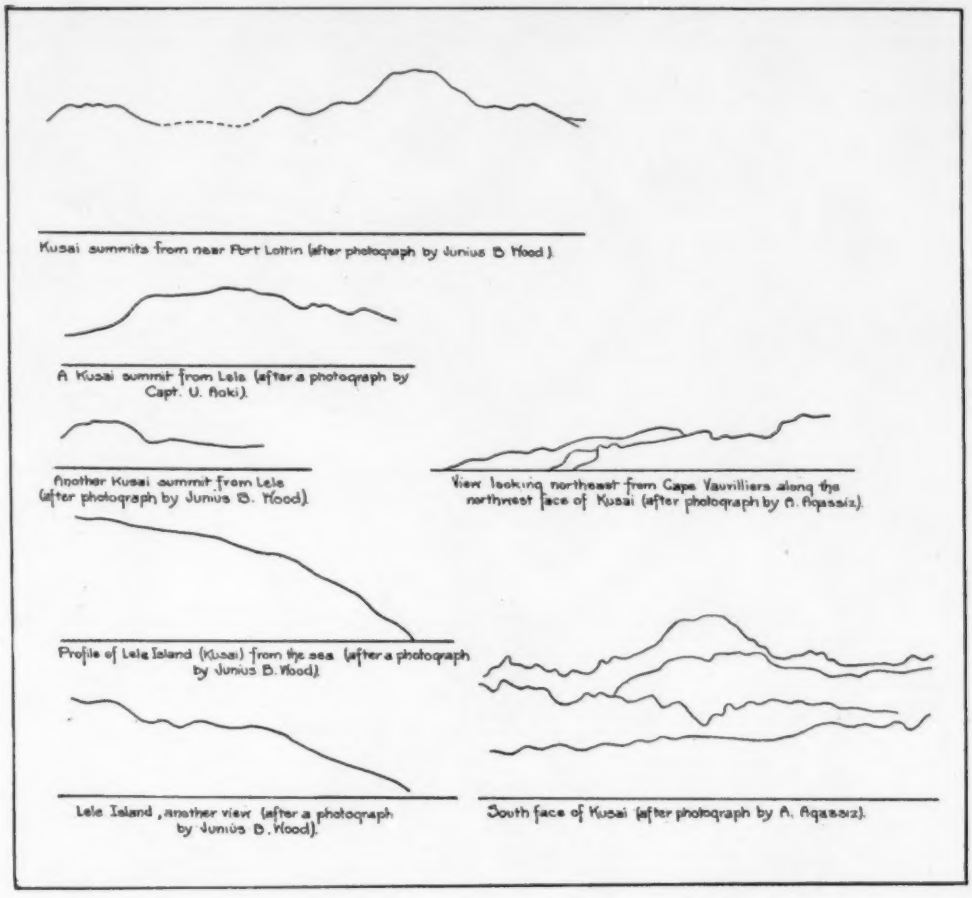

Kusai summits from near Fort Lottin (after photograph by Junius B. Wood).
A Kusai summit from Lele (after a photograph by Capt. U. Aoki).
Another Kusai summit from Lele (after photograph by Junius B. Wood).
View looking northeast from Cape Vauvilliers along the northwest face of Kusai (after photograph by A. Agassiz).
Profile of Lele Island (Kusai) from the sea (after a photograph by Junius B. Wood).
Lele Island, another view (after a photograph by Junius B. Wood).
South face of Kusai (after photograph by A. Agassiz).

FIG. 40. Profiles of Kusaie, a lava dome in the eastern Carolines; Kusai in the figure should be Kusaie.

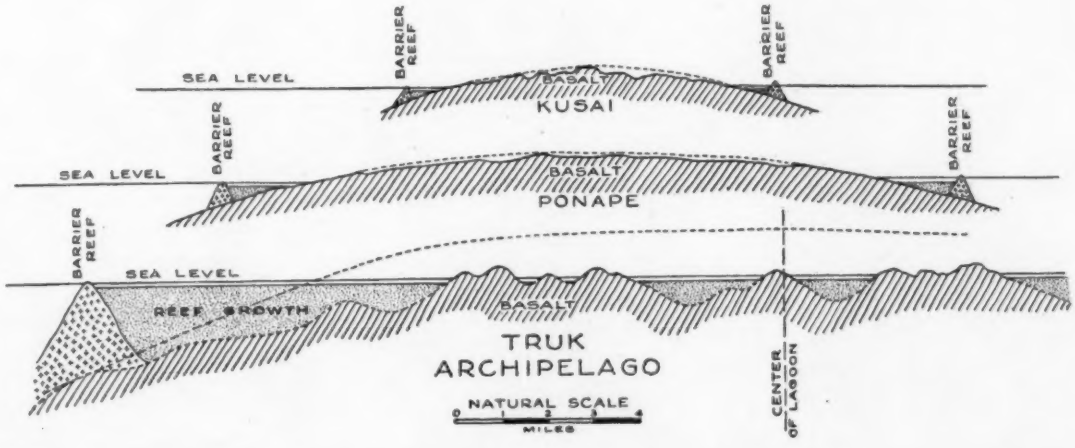


FIG. 41. Generalized profiles of the volcanic islands of the eastern Carolines; Kusai in the figure should be Kusaie.

by Hunt[60] and found to be olivine basalt, confirmed by the chemical analysis by R. K. McAlpine.

The three islands of Truk, Ponape, and Kusaie are represented by the generalized profiles on figure 41 to show their relationship.

7. SUMMARY AND CONCLUSIONS

There are two major types of mountains: (1) arcuate (folded) mountains, which form near the margins of vast subsiding areas, the basins of the oceans; and (2) extravasation mountains formed along fractures within the subsiding areas.

The arcuate mountains appear to develop where there is an initial dip in the surface of the ocean floor. The early pattern to rise is a series of arcs which are convex toward the subsiding area, from which the actuating thrusts are directed. In profile in the early stage they are made up of a flattish anticline behind a well-defined deep syncline, a trough of the sea. As the anticline evolves further, volcanoes rise at the back of the crest and exude lava of an intermediate composition, andesite—the composition of mudstone or shale. Terraces, which in tropical seas are veneered with reef limestone, are notched on the rising front of the anticline, while barrier reefs are forming on the sinking back. In later stages the arc is of sharpened curvature, and erosion mutilates more and more the model of its surface.

Within the Pacific region the youngest folded mountains compose a festooned rim which is marked by a belt of excessive seismicity and is sharply set off from the vast subsiding region at its front. This partition line, which separates mountains of the two major types in the southwest Pacific area, starts near Japan, continues east of the Marianne and Pelew Islands; then north of the Bismarck, Solomon, New Hebrides, and Fiji Islands; and, then turning southward, borders on the east the Tonga, Kermadec, and New Zealand Islands.

[60] Hobbs and Hunt, *op. cit.*

FIGURAL AFTER-EFFECTS

AN INVESTIGATION OF VISUAL PROCESSES

WOLFGANG KÖHLER AND HANS WALLACH

Department of Psychology, Swarthmore College

CONTENTS

		PAGE
	Introduction	269
I.	Preliminary experiments	271
II.	The distance paradox	281
III.	Displacements	290
IV.	Gibson's discoveries	306
V.	Direct currents and their after-effects	315
VI.	A theory of figural after-effects	327
VII.	The theory continued	341
	References	357

INTRODUCTION

THE concept of figural after-effects is just emerging from its formative stage. About ten years ago remarkable instances of such effects were discovered by J. J. Gibson. But his observations referred to quite particular figural conditions. Unfortunately an investigation of reversible figures with which we were occupied at the time, and which dealt with certain other effects of prolonged inspection, seemed equally restricted in scope. Thus it happened that for years the close relationship between Gibson's experiments and our own work was not realized. In our present report we hope to give the concept of figural after-effects a clearer connotation so that in the future such effects will be more easily recognized.

Gibson reported his first discovery in 1933 (7).[1] His subjects observed that during prolonged inspection slightly curved lines gradually became less curved. When afterwards straight lines were shown in the location and orientation of the curves, such straight lines appeared curved in the opposite direction. Their distortion could be measured. The effect was again measured by Bales and Follansbee (1), who also added some new observations. Gibson himself, however, did not restrict his experiments to the case of visual curves. He asked blind-folded subjects to move their fingers along a convexly curved edge and to repeat this movement for several minutes. They reported that the curve gradually appeared less convex. Afterwards a straight edge felt definitely concave. In vision he found that not only curves gave clear after-effects but also lines which were bent at the middle into an obtuse angle. When the apex of the angle had been fixated for some time, straight lines of the same location and orientation appeared bent in the opposite direction.[2] A further phenomenon was observed and measured by Gibson (7, 8, 10), M. D. Vernon (28), and Gibson and Radner (9). When their observers inspected a straight line which was moderately tilted with regard to the vertical or the horizontal, afterwards the objective vertical or horizontal appeared tilted in the opposite direction. Moreover, when the position of the vertical (horizontal) was thus altered the horizontal (vertical) tended to turn in the same direction.

Apart from these basic observations the authors, particularly Gibson himself, reported further facts which served to clarify the nature of such after-effects. In the *first* place it was found by Gibson that the after-effects were fairly closely restricted to the locus of the inspection figure. Here the term locus is to be understood in reference to the visual sector of the nervous system, because an effect which was caused by inspection of a figure in one place would show on an appropriate test object elsewhere as soon as the eyes were turned into the right position relative to this object. *Secondly*, however, Gibson observed that when only one eye was used during the inspection period an after-effect of somewhat smaller amount could clearly be observed in the corresponding part of the other eye's field. *Thirdly*, long inspection times, although desirable for certain purposes, proved not to be necessary for noticeable after-effects. When Gibson and Radner measured the "tilted line" effect after varying periods of inspection they found that it was unmistakably present after five seconds, and about maximal as early as one or two minutes later. With such inspection times the curve of the development tended to become parallel to the abscissa. Particularly impressive is the

[1] This and similar notations refer to the articles and books in the References at the end of the paper.

[2] Gibson (10, p. 562) mentions that this effect was first observed by F. H. Verhoeff (27).

fourth observation. Once a strong figural after-effect had been obtained it often persisted for many minutes. As a matter of fact, with one of Gibson's subjects the after-effect did not disappear within twenty-four hours when by a proper device the inspection of vertical curves had been continued for several days. But, *fifthly*, individual differences as to the amount of the after-effects were quite conspicuous, although only Bales and Follansbee found them entirely absent in some subjects.

We agree with most factual statements made by Gibson and by the authors who continued his work. But we do not believe that Gibson's interpretation of his data is correct. He assumes that it is deviation of inspection objects from "norms," like straight lines in general and verticals or horizontals in particular, which leads to figural after-effects. Actually, figural after-effects are not restricted to such special instances. It will soon be seen that, as a matter of principle, inspection of any specific entity in a visual field can cause figural after-effects. Gibson's discoveries will therefore have to be re-interpreted within the larger body of facts which is now at our disposal.

The existence of figural after-effects in a more general sense was first inferred from the behavior of reversible figures. Some data concerning the spontaneous reversals of such figures under conditions of prolonged inspection have been given elsewhere (17, pp. 67 ff.). Here we will repeat merely that the speed of those reversals tends to increase as fixation is continued, that then the figures will still be unstable after rest periods of several minutes, and that for this reason their instability can be enhanced in a sequence of separate inspection periods. This fact of summation or accumulation suggests that prolonged presence of a figure in a given location tends to operate against further presence of this figure in the same place. It may be concluded that the presence of a figure in the visual field is associated with a specific figure process in the visual sector of the brain, and that this process gradually alters the medium in which it occurs. In a reversible figure a redistribution of the figure process seems to occur when that change has reached a certain level.

It follows from these assumptions that a figure which has become unstable in its original location must again appear more stable when it is shown in a new position. Such a recovery can actually be demonstrated.

Quite apart from any reference to Gibson's discoveries, our assumptions contained certain implications which were not restricted to reversible figures. However, we did not at once realize that our interpretation involved *two* hypotheses, and that only one of them was concerned with reversible figures. According to the first hypothesis a specific figure process occurs whenever a figure appears in the visual field. And this process tends to block its own way if the figure remains for some time in the same location. The second hypothesis states that in reversible figures the figure-ground relationship will suddenly be reversed when the figure process has altered the medium beyond a critical degree. Obviously, only the second hypothesis refers to reversible figures. In the first hypothesis such particular patterns play no part. This assumption must therefore apply to all figures without exception. In other words, continued presence of *any* figure in a given location must change conditions for subsequent figure processes in the same region of the field.

When this inference was finally drawn we did not, of course, know what particular symptoms of such after-effects we had to expect in the absence of reversals. In an earlier report (17) experiments have been described which revealed some such symptoms. To be sure, these first observations gave only a very limited view of the facts. It seems nevertheless advisable to begin the present report with a discussion of the same first experiments. This will allow us to show how one group of observations after another forced us to alter our concepts considerably.

On the whole, therefore, our first chapter will repeat our previous description of certain figural after-effects. In the second chapter it will be shown that certain assumptions which we originally held were much too conservative. In the third we hope to prove that the most essential phase of figural after-effects is a withdrawal of test objects from regions which have been affected by inspection of a figure. In the fourth chapter Gibson's discoveries will be interpreted in terms of this principle.

The fifth, the sixth, and the seventh will be concerned with theory. As a preparation certain physical and physiological data will be surveyed in the fifth chapter. In the sixth and the seventh a theory of figural after-effects will be developed on those grounds. To a degree this theory will also refer to vision in general.

CHAPTER ONE

PRELIMINARY EXPERIMENTS

I

For a first observation an outline figure, e. g., an oblong or a circle, is drawn with india ink on a white screen, and a fixation mark is added on one side (*cf.* Fig. 1). The outline ought not to be too thin. On a second white screen the same drawing is prepared in the same position relative to the fixation mark, but on the other side of this mark, and symmetrically, an identical figure is here added.[3] The subject first fixates the mark on the former chart for, say, five minutes and from a distance of two yards. If he then fixates the mark on the second chart he will find that its two objectively equal figures have not the same appearance. The figure which coincides with the previously inspected object will be pale or gray in comparison with its black partner, it will seem to lie farther back in space, and it may look a trifle smaller. Depending upon the individual subject one or another of these facts will appear more conspicuous. For instance, some subjects report at once that the two objects have not the same size, while others do not mention this difference spontaneously. A few may even deny its existence. More convincing results are generally obtained with oblongs than with circles.

In a given and constant location a dark area, and therefore also a black line, may gradually assume a gray appearance by local "fatigue" rather than by a figural after-effect. If such a banal explanation can be given for one of the three symptoms one might be tempted to interpret the other two as secondary consequences of the first. They, too, would then become unimportant. However, such an explanation is quite untenable in many cases which will soon be discussed, and it is not plausible even in the present instance. For, if during the inspection period one eye is prevented from seeing the figure, and if afterwards the test is made with this eye alone, the same symptoms will be observed.

The inspection object of Figure 1 may be replaced by a solid figure, for instance, by a black solid circle. If the test objects are two equal and symmetrically situated outline circles of which one coincides with the contour of the solid figure, these test objects will differ after the inspection period more or less as they did in the first experiment. In this observation it is almost necessary to use one eye during the inspection period and the other in the test. Otherwise

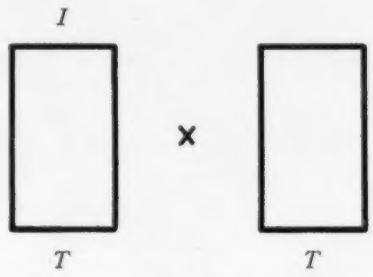


FIG. 1

the negative after-image of the solid figure would disturb the comparison. With transfer from one eye to the other few observers mention after-images, and without this disturbing factor the difference in the appearance of the test objects can be more easily seen.

Other figural after-effects are incomparably more impressive than the ones just mentioned. But it seems natural to begin with a description of these observations because they were the first we made and would probably be everybody's first choice: The "figure" of an outline object appears characterized first of all by its outline, the only objective datum which interrupts the homogeneous ground. Hence, if a figure causes an after-effect, ought this effect not to be strongest at the place of the outline, and therefore with a congruent test object?

Actually it will soon be seen that observations do not generally confirm this expectation. Before we proceed to further experiments it will, however, be advisable to make a few remarks about the simple technique which has just been used. This technique will remain useful throughout our investigation; but its application requires a certain amount of caution.

In many experiments one test object which lies within the affected region or in its immediate neighborhood will be compared with an equal second test object which is placed in a neutral region. Whether this latter region is entirely unaffected is not a very important question. So long as under favorable conditions a striking difference between the test objects is seen the procedure will serve our purpose. It is of course

[3] Our Figures show inspection objects (*I*) and test objects (*T*) in one pattern so that their relative positions can readily be seen. When, as in Figure 1, a test object coincides with the inspection object, this fact is indicated by the letters *IT*. Wherever no particular dimensions are given our actual Figures were linearly about 30 per cent larger than the present reproductions.

assumed that at a *great* distance from a previously inspected object after-effects are smaller than they are in its neighborhood. But this assumption is corroborated not only by Gibson's observations but also by our own experiences. For the most part we have shown one test object in the left half of the visual field and the other in the right half. We recommend the use of test patterns which are symmetrically placed with regard to the main axes of space. Under these conditions intrinsic inhomogeneities of the visual field play but a minor role. Moreover, in a symmetrical situation the comparison of test objects will be found to be easier.

More dangerous than a possible spread of after-effects into the region of the neutral test object are other influences upon this region which one is likely to overlook in the beginning. A concrete example: In observations such as the ones just described it would be simpler to use not two charts but only the one which contains the two test objects. One of these could first be covered by a screen and the other thus used as the inspection object. After the inspection period the two objects could be directly compared when the screen is removed. Although we have sometimes used this procedure we cannot recommend it. The covering screen itself constitutes a figure; at least one of its edges will lie within the field, and after-effects will result which may influence the outcome of the experiment. As a matter of fact, the after-effect of the screen may sometimes be stronger than the effect which a given experiment is meant to demonstrate.

While with some figural after-effects a small fraction of a minute will suffice as an inspection period others require a longer inspection time, particularly if the subjects are not accustomed to such observations. This often makes experimentation distinctly unpleasant. Observers complain about headaches, also about a feeling of strain and of disagreeable dryness in the eyes. We are not at all sure that the investigation of figural after-effects belongs to the occupations which an oculist or a neurologist would recommend. This suspicion is enhanced by the fact that once an after-effect has been well established it may not disappear for hours. We quite agree with Gibson's remarks on this point.

However relevant to any theory of figural after-effects their long survival may be, it constitutes unfortunately an important obstacle to reliable experimentation. For the most part experiments on figural after-effects will have to be separated by considerable periods of intermission. Otherwise an older effect will easily falsify the outcome of a subsequent observation. Before any experiment is begun the subject will have to make sure that the appearance of the test objects is normal. This is in any case a wise precaution because objectively equal objects may not be visually equal when shown in different locations. The astonishing persistency of certain after-effects makes the interpolation of considerable rest periods a necessary condition of most experimentation in this field.

When two charts are being used, one during the inspection period and one for the test objects, the latter may be shown in the place which was first occupied by the former. Sometimes it will be more convenient to arrange the two charts in symmetrical positions, one to the left and the other to the right of the subject's median plane, and with a slight inclination toward this plane. A small and symmetrical turn of the head will then allow the subject to fixate the mark first on one and then on the other chart as though they were shown in the same place and both right in front. However, this procedure is admissible only so long as the subject avoids any tilting of his head while turning from one chart to the other. Any tilt would obviously alter the spatial relation between the affected area and the corresponding test object.

Few observers are disturbed by negative after-images when the inspection object consists either of black lines on a large white ground or of white lines on a dark ground. With solid figures, we remember, the disturbance by after-images can be avoided if the inspection object is shown only to one eye and the test objects only to the other.

We suspect that under these circumstances transfer of the effect cannot simply be taken for granted. At least its amount may depend on the conditions under which an experiment is being performed. It seems to us that transfer is more complete if during the inspection period the second eye is not closed or covered, i.e. darkened, but merely prevented from seeing the inspection object. A screen which is held obliquely before this eye will keep it illuminated and at the same time cover the inspection object. In this position edges of the screen will be so dimly seen that they can hardly act as effective contours.

II

The following experiments differ in two respects from those of section I: The test objects are no longer equal to the inspection objects, and consequently the outline of the former does not, or does not throughout, coincide with that of the latter. It will be seen that this change does not obliterate the after-effects; on the contrary, the next observations will generally be regarded as more convincing than those of section I.

1. In Figure 2 the inspection object is an outline circle. The subject fixates a mark on the

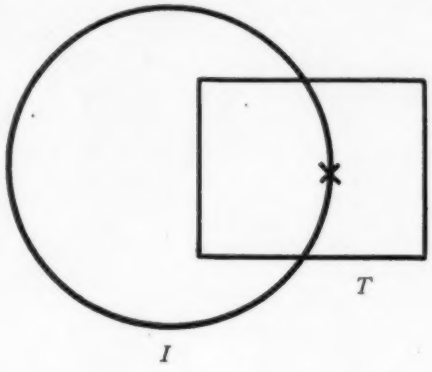


FIG. 2

periphery of the circle. The test object is an oblong divided by a curve into approximately equal parts. This curve is congruent with the corresponding part of the circle and bears a fixation mark where that of the circle lies. In Figure 2 the inspection object (*I*-object) and the test object (*T*-object) are united; actually the *T*-object is shown when the *I*-object is no longer visible. After an inspection period of several minutes all subjects report that the right side of the *T*-oblong is larger, particularly higher, than its left side. It will surprise every observer to find that after a sufficient inspection period the horizontal contours of the oblong are actually

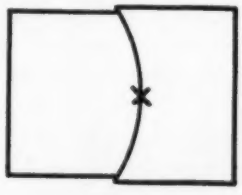


FIG. 3

broken so that it assumes the shape indicated in Figure 3. A depth effect is no less obvious. Normally the convex part of the oblong tends to appear in front of the concave part. After the inspection period just the opposite is true; the concave part now lies "upon" the convex half; in other words, the figure-ground relationship is reversed.[4] Obviously in the interior of the *I*-figure the figural after-effect is stronger than it is immediately outside its contour.

The *I*-object need by no means be a circle. When the observation is repeated with a rectangle and therefore with a *T*-object divided by a straight line it gives precisely the same result.

2. When experiment 1 of this section was first performed the outside part of the *T*-object appeared so "intense" that for a moment the following interpretation seemed possible. Inside the figure area prolonged inspection causes an effect of "depression" while just outside the figure the opposite holds. Actually this is not the case. When a *T*-object just outside an *I*-figure is compared with a second one which lies at a great distance, the former will be found to show all symptoms of "depression." This can be demonstrated with the arrangement shown in Figure 4.

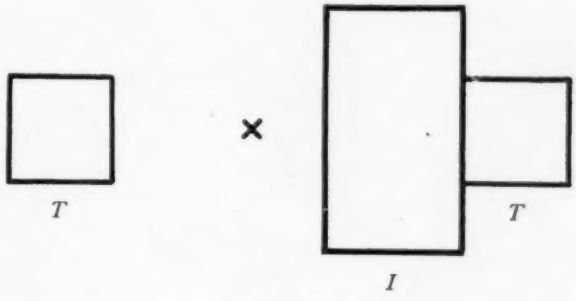


FIG. 4

(In this figure the *I*- and the *T*-objects are again united while actually they are shown in succession.[5]) After the inspection period the *T*-square on the right side will look considerably smaller; it will also lie farther back in space and be pale in comparison with its partner. It follows that the after-effect extends beyond the area of the inspection object, and that its symptoms are here the same as they are within the figure. In ex-

[4] It will be realized that this observation establishes a direct connection between the present after-effect and the behavior of reversible figures. Actually the *T*-object of experiment 1 may be regarded as a reversible figure.

[5] In the following pages this interpretation of our Figures will no longer be mentioned. Parts which are marked with an *I* are invariably inspection objects, and are shown first; parts designated as *T* are test objects, and shown after the inspection period. The fixation mark (×) is common to both.

periment 1, therefore, the outside part of the *T*-object seemed particularly "intense" only in comparison with the more strongly affected inside part.

3. In a further experiment a figural after-effect is demonstrated when the outline of the affected *T*-object has no point in common with the outline of the *I*-figure (Fig. 5). After an inspection

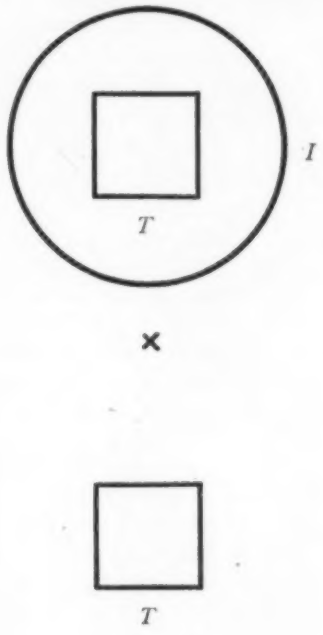


FIG. 5

period of several minutes no subject will miss the striking difference in the appearance of the *T*-squares. The square above the fixation point will appear smaller than the other *T*-object; it will also lie farther back in space. Under the conditions of this experiment, therefore, the after-effect pervades the whole interior of the *I*-object.

Within limits the particular nature of the *T*-object may be varied. If the squares are replaced by pairs of horizontal lines, one pair within the circle, the other symmetrically below the fixation point, the after-effect remains just as striking. Within the area of the circle the *T*-lines appear gray, not black; they lie back in space, and are probably nearer one another; they are also clearly shorter.

4. There are no figures without outlines or contours. It is, therefore, a plausible assumption that in some way contours are responsible not only for the existence of figures as such but also for any after-effects which are established when figures are shown for some time in a constant location. We should not conclude, however, that during prolonged inspection of a figure the various parts of its outline emit, independently of one another, some detrimental agent which causes a "depressing" effect wherever it arrives.[6] From this point of view an after-effect ought to become stronger in proportion to the amount of contour—of an *I*-object—which surrounds the area of the *T*-object at a given distance. We shall see in Chapter Two that whole classes of after-effects are incompatible with this view. For the moment, the outcome of a simple "differential" experiment will suffice as a refutation. In Figure 6 the broken lines of the *I*-square

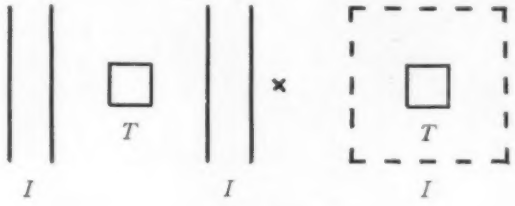


FIG. 6

on the right side of the fixation mark "surround" the right *T*-object more completely than the two inner verticals surround the left *T*-object. The distance of these verticals from each other has the same length as the edge of the *I*-square on the right side. But in terms of actual contour each edge of the square is only two-fifths of each continuous vertical on the left side. The total actual contour of the square is, therefore, in this measure eight-fifths while the length of the two continuous lines is two, or ten-fifths. In other words, the total length of actual contour is twenty-five per cent greater in the case of the two continuous lines. Moreover, on the left side two more verticals are added on both sides of the *T*-object. Distances of the *I*-contours from the *T*-objects are in this pattern directly comparable, because the broken lines of the square have the same location relative to the right *T*-object as the inner continuous *I*-lines have to the left *T*-object. Thus the experiment

[6] This assumption has been made by K. J. W. Craik and O. L. Zangwill (3) in their interpretation of experiments in which surrounding outlines were shown to raise the threshold for the appearance of a weak dot. Incidentally, during these experiments tremendous after-effects must have developed. It remains to be seen to what extent these after-effects explain the results which the authors contrast with a theoretical prediction of Koffka's.

can decide whether, apart from the amount of contours in the neighborhood of a *T*-object, the particular configuration of these contours plays a part in determining the after-effect.

The answer is perfectly clear. The *T*-object on the left side appears larger and, for most observers, also nearer. It follows that the strength of an after-effect depends upon the specific configuration in which the *I*-contours are given in the neighborhood of a *T*-object. Mere quantity of contour at a given distance is surely not the only decisive factor.

5. Experiments 1, 3, and 4 have proved that the interior of a figure is in an altered condition when this figure has been for some time in the same place. It seems, however, equally relevant to know to what extent the effect extends into the environment of a figure. In experiment 2 we have found that a *T*-object which lies just outside the figure is clearly "depressed." The next experiment shows that this influence of the *I*-figure is not restricted to *T*-objects in an immediately adjacent position.

Figure 7 is to be interpreted as follows: The *I*-object, a solid black oblong, is shown with the

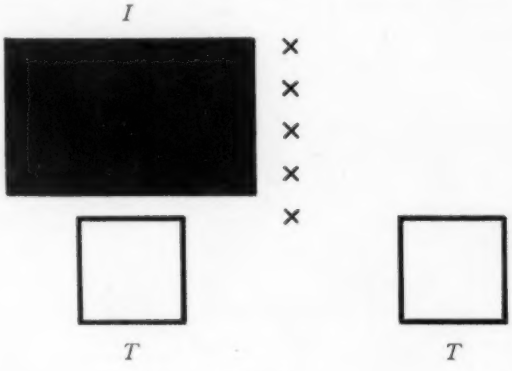


FIG. 7

lowest of the five crosses as a fixation mark. After prolonged inspection the two *T*-squares are compared when the eyes fixate either the same or one of the higher marks. (These higher marks are given only on the *T*-chart, not during the inspection period.) The higher the mark lies which is fixated in the test the greater will be the distance between the area previously occupied by the *I*-object and the left *T*-square.[7] Thus a number of tests with varying distances can be made with the same *T*-chart. Short inspection periods between successive tests will keep the after-effect at a high level if inspection has been sufficiently long in the beginning. Our subjects observed from a distance of about two yards, a distance which was used in most of our experiments.

All observers reported that the left *T*-object was affected when they fixated the first, the second, or the third mark from below. With some subjects the limit was not reached at this point. Since the fixation marks were two-fifths of an inch distant from one another, and the first one-fifth of an inch distant from the nearest contour of the *I*-object, this observation means that in the present experiment the after-effect extended for all subjects at least one inch beyond the area of the *I*-figure. Again we found variations as to the particular symptom which seemed to impress individual observers most. One subject would stress the comparative paleness of the left *T*-object, another its smaller size, again another the fact that it lay back in space. Frequently a fourth symptom was mentioned, a displacement of the left *T*-square. We prefer to discuss this new fact in another section.

As preliminary evidence this observation will suffice. Although it is quite true that figural after-effects are centered about the locus of the *I*-figure they are by no means strictly limited to this place. Therefore it does not seem entirely relevant to ask whether or not these effects are "localized." Their center has a very definite location; but from this center the temporary disturbance extends considerably into the environment.

From this fact we can draw an important conclusion. It will not be denied that the prolonged presence of a figure in a given area is the cause of the after-effects. This can mean only that continued occurrence of a specific figure process in a given location changes the state of the medium in which this process occurs; that in the changed medium the figure processes of *T*-objects are altered; and that as a consequence the visual characteristics of these *T*-objects are affected. But from this point of view the extension of the after-effects in space will roughly correspond to the extension of the figure processes by which they are established. It follows that, although the color processes of a given figure are limited to the area of the figure as such, in some way the

[7] The same holds for the square in the right half of the field. Therefore, when successively higher fixation points are used the difference between the distances of the squares from the area of the *I*-object becomes relatively smaller. With the largest distances this may tend to weaken the observed after-effect.

figure must act beyond this restricted area. Otherwise no after-effects could be observed outside and at a distance from the *I*-figure. In this sense, then, a figure has a "field" by which it is represented in the environment.

Our conclusion that a figure is surrounded by a field may seem to be open to the following criticism. The characteristics of a *T*-object need not depend merely on the state of the medium at the locus of that *T*-object. To a degree they may depend on alterations of the medium in other places, for instance within the area of the *I*-object. This criticism defeats itself in assuming that the *T*-object is associated with events which extend beyond the area of the *T*-object. How else should the *T*-object be affected by conditions at a distance? But this is precisely the assumption which we have made with regard to the *I*-object. Thus the argument implicitly grants the thesis against which it is aimed. It should, however, be admitted that because of the field of the *T*-object our present observations give at best a very rough indication of the extension of the *I*-field.

At this point a terminological remark becomes necessary. We have to distinguish between two facts. A figure process, it appears, tends to alter conditions within the medium in which it occurs, and when the figure disappears the medium seems for some time to remain in its changed condition. This is a state of affairs for which, apart from the medium itself, only the figure process of the *I*-object is responsible. Suppose now that a *T*-figure is shown in the affected region. Generally speaking, not all parts of this region will be equally affected. Consequently any alterations which the process of the *T*-figure may suffer will depend upon the particular position which this *T*-figure has in the affected zone. Furthermore, the particular size and shape of the *T*-figure will also influence such after-effects. Hence, within a region which has been affected by inspection of a given figure not one but several after-effects will be demonstrable. Under these circumstances it would be confusing if both the alteration of the medium as such and the various effects of this alteration upon specific *T*-objects were given the same name. We propose to call only the alterations of *T*-objects "figural after-effects" and to refer to the affection of the medium as "satiation." This name is not meant to have any particular implications beyond the fact that the prolonged presence of a given figure causes the "depressed" condition of the medium. As soon as the nature of this depression becomes sufficiently known the term satiation may of course be replaced by a better defined concept.

III

The observations of sections I and II refer to a few particular patterns. From this one should not infer that inspection only of special objects causes satiation and figural after-effects. It has been shown that both solid and outline figures can be used. But we obtained equally strong effects with grid-like patterns. Furthermore, a simple straight line suffices as an *I*-object. In the arrangement of Figure 8 the two squares are

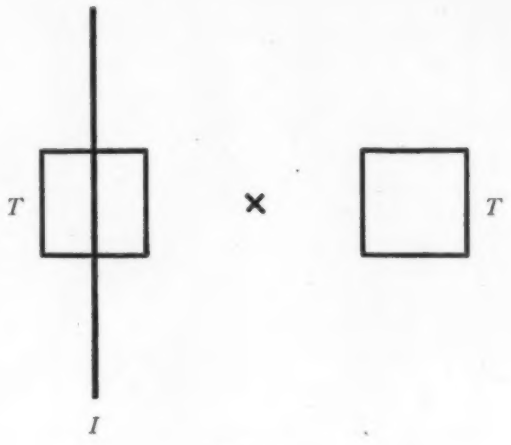


FIG. 8

T-objects. After inspection of the line for several minutes the left square will show the now familiar symptoms of depression. The direction of the line and correspondingly the location of the *T*-squares may of course be varied. Again, the effects are not restricted to the locus of the line itself. Although less conspicuous, they will still appear in arrangements like those of Figure 9

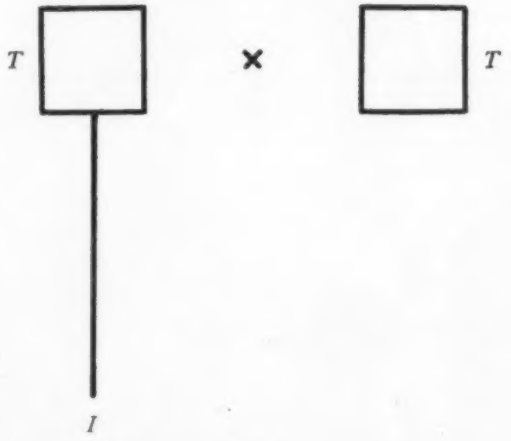


FIG. 9

and Figure 10. In Figure 10, for instance, the square on the right side will be found to be comparatively pale, to lie farther back in space, and to be smaller. The difference as to size is likely to be particularly pronounced in the dimension parallel to the straight line.

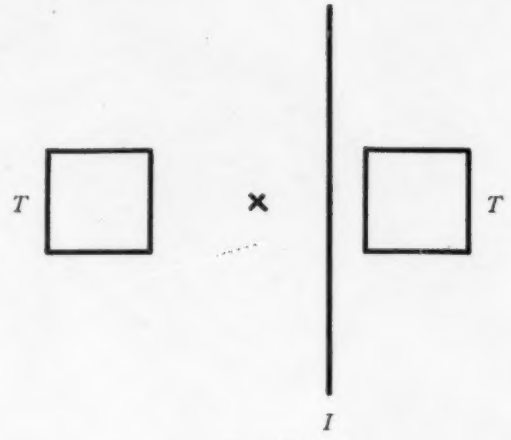


FIG. 10

We cannot reconcile this observation with Gibson's view that figural after-effects are caused solely by deviations of *I*-figures from certain norms. To be sure, on this ground none of the effects described in sections I and II could have been predicted. But the fact that, after inspection of a perfectly straight line, *T*-objects are also affected actually argues against that principle. For, the straight line is one of the norms to which Gibson's principle refers. And yet, when such a line is inspected, i. e., in the absence of any deviation from the norm, after-effects are still observable.

Even a simple straight line is not the "weakest" object the inspection of which causes after-effects. Observers who are particularly sensitive to changes in the appearance of *T*-objects will sometimes report that in a very small region around the fixation point tiny *T*-figures are clearly affected. It is fortunate that these after-effects of the fixation mark do not seem to reach very far in space. Otherwise it would be difficult to separate their ever-present influence from other effects which a given arrangement is meant to test.

If a mere point is the weakest *I*-object with which we are acquainted solid things in our everyday environment are probably the most robust "figures" to be used in such experiments. The after-effects of objects in this sense do not seem to differ from those of figures in a more customary meaning of the term. As an example we once put a common sponge on a white surface and fixated a point on one side of this object. When after removal of the sponge two squares were shown in a symmetrical position with regard to the fixation mark, the square which lay in the place of the sponge appeared strongly affected. A tobacco pouch, a dollar note, the subject's own hand, or any other object which is clearly segregated from a uniform ground will serve the same purpose. In other words, every specific entity in a visual field causes satiation and, upon introduction of suitable *T*-objects, corresponding figural after-effects.[8] It is for this reason that we cannot recommend the use of any screens which the experimenter will often be tempted to use when he wishes to examine the influence of only part of a given *I*-pattern, and therefore to make other parts temporarily invisible (*cf.* p. 272). As a rule *T*-objects will be strongly affected within the area of such a screen. On the other hand, we have never been disturbed by any figural after-effects which our inspection charts, large rectangles of cardboard, might be expected to cause as figures. Such effects must be extremely weak since in our experiments both the *I*-figures and the *T*-objects are far removed from the edges of the large charts. Moreover, the location of the *T*-objects invariably approximates symmetry with regard to those edges so that any after-effects of the charts would be the same on both sides.

Figural after-effects, we have seen, are no less revealing as to the nature of the figure process *per se* than they are interesting on their own account. With the following examples we wish to demonstrate the sensitivity of figural after-effects to apparently minor changes of the *I*-pattern, and therefore of the figure process.

Figure 11 needs no explanation. The outline of the lower circle is interrupted on one side; the angular size of the opening amounts to 60°. We found that after the inspection period the lower *T*-square appeared nearer and just a trifle larger than its partner.

Even objectively equal *I*-figures may cause slightly different after-effects if their spatial ori-

[8] In passing it may be remarked that if with four figures two equal solids are formed in a stereoscope—one for prolonged inspection, the other for later use as a comparison object—after the inspection period the former will appear quite flat in comparison with the latter.

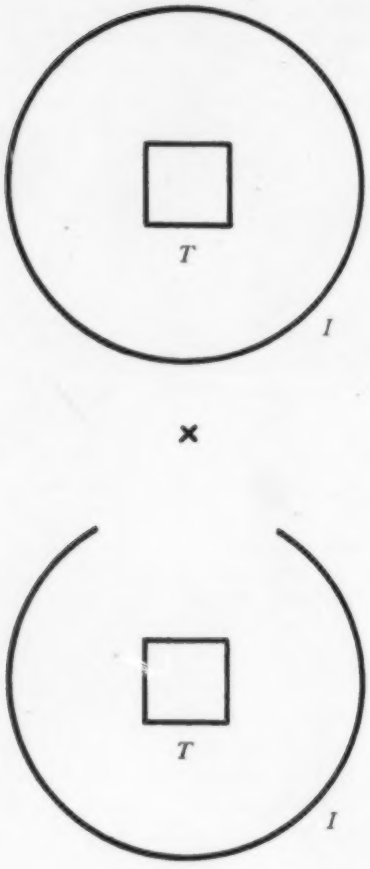


FIG. 11

entation is varied. In Figure 12 the pairs of parallel lines on both sides of the fixation mark are identical. Nevertheless their after-effects were found to be perceptibly different: the left square lay back in space. It seems to follow

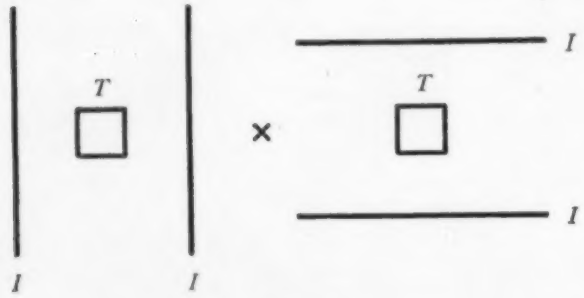


FIG. 12

that the figure process between the vertical parallels is not quite the same as that between the horizontals, that it is more intense between the verticals. Our pattern represents a special instance of the so-called Vertical-Horizontal illusion. Visually the vertical lines are longer than the horizontal lines. For the same reason the horizontal distance between the vertical lines is visually shorter than the vertical distance between the horizontal lines. It remains to be seen whether these facts are related to the present observation.

We subjected a second so-called illusion to the same procedure. Figure 13 shows the Müller-

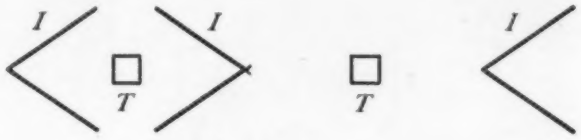


FIG. 13

Lyer pattern together with two *T*-objects which were presented after prolonged fixation of the mark in the center. Our subjects agreed that in this instance the observation was particularly convincing. The left square lay clearly behind the right square, and the former was smaller than the latter. This result was to be expected if the intensity of the figure process, and therefore satiation, was greater in the more closed area on the left side. But it seems curious that just the part of the figure which contains the stronger process appears as shorter.

The experiments represented by Figures 11 to 13 are differential experiments in which two figural after-effects are compared. In Figure 12 the difference of conditions on one side and the other is obviously slight. Under these circumstances not all observers may be able to confirm such a differential effect. With subjects whose after-effects are generally small, some differential effects may easily become subliminal. If this happens no judgment about figural after-effects as such will be in order. We shall soon turn to particularly strong effects which nobody is likely to miss. We suggest that in case of any doubt such instances be first of all examined.

Just as prolonged inspection of countless objects causes satiation and figural after-effects, so the *T*-objects need not be squares or other very simple figures. It seems particularly important that in strongly affected areas *distances* between such objects are changed just as are the sizes of the objects themselves. If a *T*-object lies within an affected zone it tends to appear smaller. Similarly, if two *T*-objects are shown in such a region they will lie nearer each other than they do in a less affected region.

In Figure 14 the oblong is the *I*-object. The four squares are the *T*-objects. After inspection

of the oblong the left *T*-squares will show the usual symptoms of depression; they will appear to lie back in space, to be paler and smaller. But quite as striking will be the fact that the left side of the *T*-pattern is shortened in the vertical

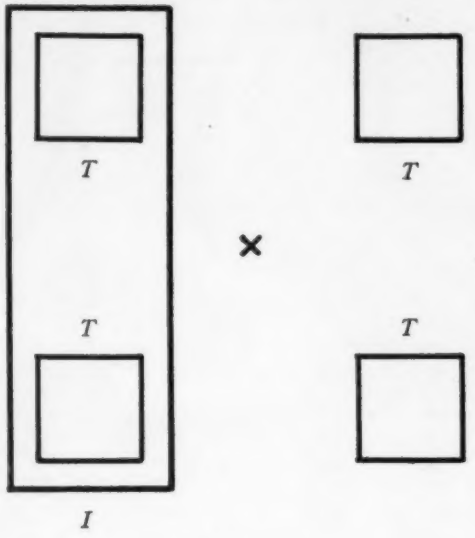


FIG. 14

direction. In other words, the *T*-pattern as a whole will appear as a trapezoid, not as a square or an oblong.

Is this influence upon the distance between two *T*-objects a secondary consequence of changes which the size of these objects themselves suffers? One might be inclined to believe that as these objects become smaller the over-all extension of the pair in the vertical direction must decrease,

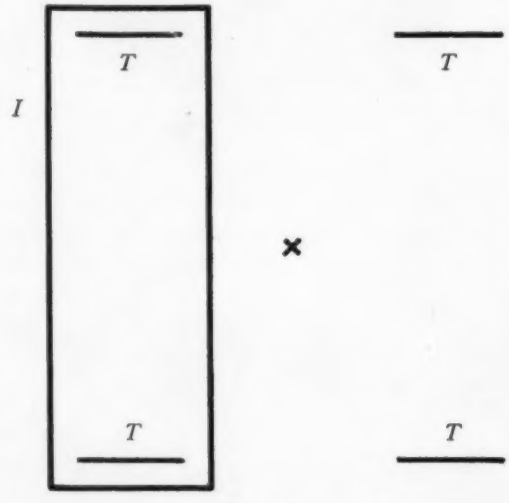


FIG. 15

and that it is this decrease which appears as a reduction of their distance. That more than this is involved can be shown with the arrangement of Figure 15. Here the *T*-objects are four thin horizontal lines the width of which cannot be appreciably altered when they lie within an affected area. None the less after inspection of the oblong the distance between the lines on the affected side is reduced. In this respect the free zone between two separate *T*-objects does not seem to differ essentially from the objectively empty interior of a closed outline object. Just as this area of a closed *T*-object appears smaller when the object is shown within an affected region, so an area which is only partly enclosed by two *T*-objects will shrink under the same circumstances.

In the further development of our investigation *T*-patterns like the four squares of Figure 14

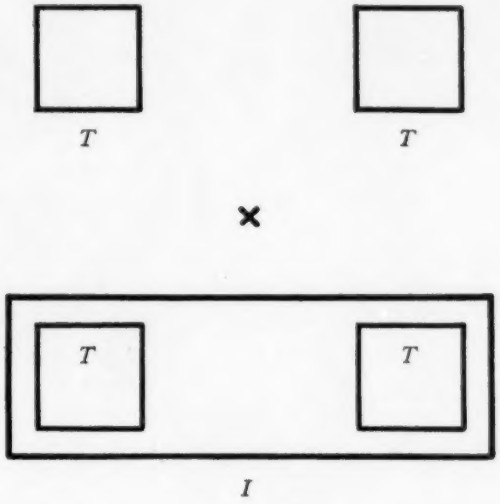


FIG. 16

proved to be extremely useful. Figural after-effects of many kinds can be tested with the same simple arrangement. We may add a few more instances which demonstrate the decrease of distances between separate *T*-objects.

In the first place, the effect is the same if we test a *horizontal* distance. After inspection of the oblong in Figure 16 the pattern of squares is no longer symmetrical; its lower horizontal distance is shortened. Secondly, if the *I*-object itself consists of two separate objects as it does in Figure 17, *T*-objects which are presented within the areas of these *I*-figures appear again too near one another. On the face of it the present effect does not seem to differ very much from the one

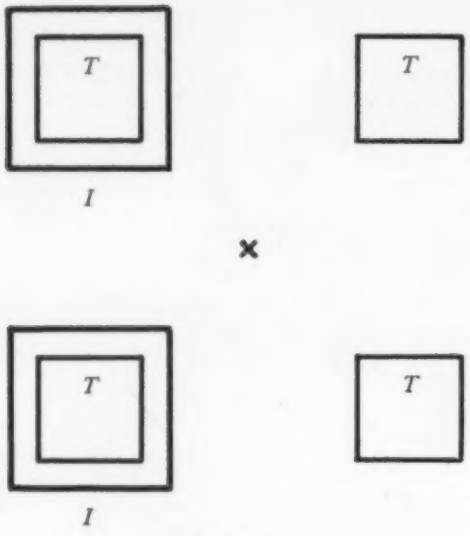


FIG. 17

in which two *T*-squares lie within the area of one *I*-object.

Again with the same *T*-pattern we can show that within an *I*-object the alteration of the medium need by no means be uniform. After inspection of the large square in Figure 18 the

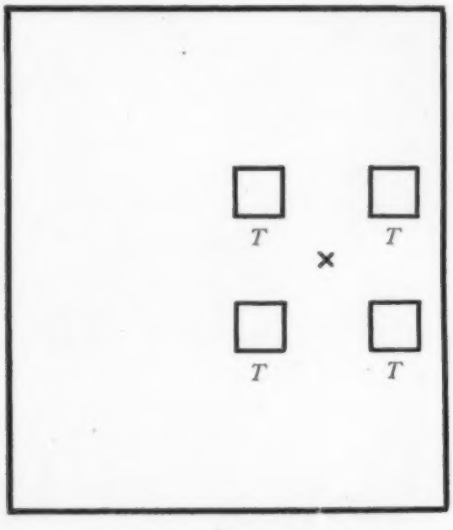


FIG. 18

four *T*-objects do not appear as a symmetrical pattern; the vertical distance on the right side, which lies nearer the contour of the *I*-object, is clearly shorter. It follows that within a sizable *I*-object satiation is more intense in the neighborhood of the contour than it is in more central regions.

Distances between *T*-objects decrease also if these objects lie in a strongly affected region *outside* the area of a previously inspected figure. After inspection of the solid oblong in Figure 19

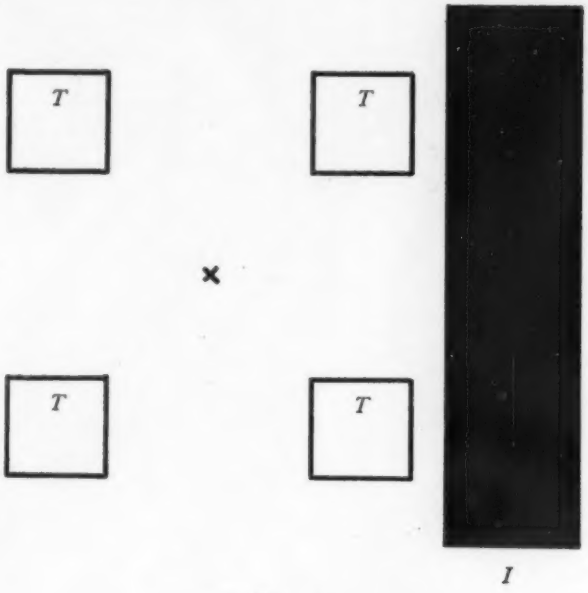


FIG. 19

the vertical distance of the *T*-objects on the right side appears much shorter than the distance of those on the left side. This effect, too, will be found to extend considerably into the environ-

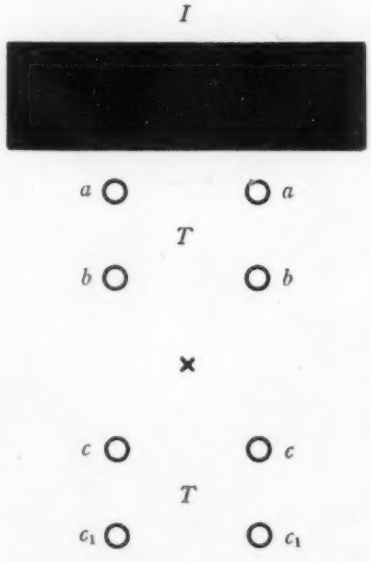


FIG. 20

ment. In the arrangement of Figure 20 the distances *cc* and c_1c_1 remain equal. At such distances from the *I*-figure satiation must either be absent or have but a weak gradient. On the other hand, the distance *bb* is clearly shortened although less so than *aa*. This confirms our observations on p. 275; because in the present pattern the distance between the *b*-objects and the edge of the oblong is not much more than one inch.

Nobody will operate for long with such *T*-patterns without being surprised by a curious phenomenon. In the first experiment in which the pattern of four squares was used (*cf.* p. 279) the two squares on the left side appeared nearer each other than the squares on the right side because the former fell into a strongly affected area. But Dr. R. Crutchfield who was then one of our observers remarked at once that this distortion of the *T*-pattern was not symmetrical: the distance on the left side was shortened more by a displacement upward of the lower *T*-object than by an opposite change in the position of the higher object. Ever since this first observation the same fact has impressed us in practically all experiments in which the nature of the test permitted its occurrence. With regard to satiation, just as in other respects, the visual field clearly constitutes an anisotropic medium. Not only are figural after-effects generally stronger in the lower half of the field, they also appear here earlier. After a very short inspection period, therefore, subjects may report distortions only in the lower half of the *T*-pattern. We shall soon find that the same rule holds for patterns and after-effects which have otherwise little resemblance with the present observation. For instance, Gibson's "curved line" effect often shows only in the lower half of the *T*-line when inspection has been restricted to ten or fifteen seconds. On the other hand, no such asymmetry seems to exist in the horizontal dimension. If within an evenly affected area a horizontal distance is shortened this distortion is to our knowledge always symmetrical. So it appears at least if the objective pattern itself is placed symmetrically with regard to the median vertical.

Obviously, the asymmetry of figural after-effects is unrelated to the nature of the patterns which serve as *I*-figures and as *T*-objects. These objects may be perfectly symmetrical above and below the equator of the visual field; and yet their lower parts will be more easily and more strongly affected. This can mean only that the medium in which visual figure processes occur is anisotropic in the vertical dimension. Its anisotropy must cause an asymmetrical distribution of those processes, or an asymmetrical satiation, or both asymmetries at the same time.

CHAPTER TWO

THE DISTANCE PARADOX

I

The experiments described in Chapter One have been repeated so often and with so many subjects that, as facts, they deserve full confidence. If we were to believe, however, that our report provides an adequate knowledge of figural after-effects we should be mistaken. The next two chapters will show that an essential phase of most figural after-effects has not been mentioned in the descriptions which we have so far given.

Two reasons made us suspect that an important point had still escaped our attention. First, a certain aspect of our own observations seemed to us paradoxical. Second, we could not see a clear relation between figural after-effects as here described and figural after-effects as discovered by Gibson. In the present section we propose to make the paradoxical nature of our findings quite explicit. This will facilitate realization of the crucial fact in our field which we have not yet recognized. Once this fact is known we shall have no difficulty in connecting Gibson's experiments with our own observations.

With the possible exception of some differential experiments the observations reported in sections II and III of the first chapter are not likely to offer serious difficulties to those who wish to repeat them. Some after-effects will be found to be just as striking as a number of well-known "illusions" in which patterns appear distorted at first sight.

The same cannot be said of the experiments which were described in section I of Chapter One. When a *T*-object coincides with the *I*-object, the comparison with an equal, second, *T*-object in an unaffected area will not always give entirely satisfactory results. Two symptoms, it is true, may be expected to be clear enough: if the *T*-object is a black outline figure which coincides with the *I*-object it will be pale; and it will probably lie farther back in space than its unaffected

partner. But of the third symptom, a shrinking of the *T*-object, we can under these circumstances not be so sure. And yet, in the experiments of sections II and III changes of sizes (and distances) are particularly impressive. Take, for instance, experiment 3 in section II. We have not yet found a subject who would fail to observe that the *T*-object which lies within the area of the *I*-circle is smaller than its unaffected partner. Why, then, is the shrinking of a *T*-object more conspicuous when this object lies in the interior of the *I*-object than it is when both coincide?

In showing a *T*-object within an affected area we have tacitly proceeded as though the visual field were comparable to a medium in which the temperature has been locally raised or lowered, and as though the function of the *T*-object itself were that of a thermometer which measures such local changes. In other words, it has been supposed that when exhibiting a figural after-effect the *T*-object indicates the degree of satiation at the place at which the object is shown.

So far as changes of color and of position in the third dimension are concerned this interpretation of figural after-effects may be entirely correct. On the other hand, when applied to the third symptom, the decrease of the size of *T*-objects, this view of figural after-effects proves to be incompatible with numerous facts. The first evidence to militate against it is our observation that within the area of an *I*-figure the size of a *T*-object is more strongly affected than it is when the *T*-object coincides with the *I*-object. For, whatever our theory of figure processes may be we shall always be forced to assume that between the occurrence of these processes and the presence of contours and outlines there exists a direct relationship. It seems to follow that figure processes are concentrated about, or that they issue from, the outlines of their figures. On this premise, since satiation is caused by the figure process, satiation ought to be maximal in the region of the outline. Actually, however, in this location the size of a *T*-object seems to be much less reduced than is that of a smaller test-object which, one should think, lies in a less affected area. Therefore, either our assumption about the spatial distribution of figure processes and satiation is wrong or our interpretation of this particular symptom, the change of size of a *T*-object, is in need of correction. Everything points in the direction of the second alternative: a *T*-object which is shown in a certain location is not in every respect comparable to a thermometer, and a figural after-effect which it may exhibit does not merely indicate to what extent the medium has been satiated in that location.

To prove this point we had first of all to make sure that the factual basis of our argument was sound. We decided to measure the effect which previous inspection of a larger outline circle has upon the size of a smaller circle in its interior. Both the *I*-circle and this *T*-circle were shown on the left side of the fixation mark, and the position of the *T*-object was concentric with regard to the area previously occupied by the *I*-circle. On the right side of the field circles of varying diameter could be shown as neutral *T*-objects, and the subject compared in each case the size of the *T*-object in the affected area with that of its partner (*cf.* Fig. 21). Actually the *I*-object was drawn on one chart, and the various pairs of *T*-objects on a series of further charts. The outline of the *I*-circle was $\frac{1}{8}$ inch thick; its outside contour had a diameter of $5\frac{1}{16}$,

FIG. 21

its inside contour of $4\frac{13}{16}$ inches. The *T*-circles were not more strongly drawn than usual. The (affected) *T*-circle on the left side had a constant diameter of $4\frac{2}{8}$ inches while the diameter of the comparison circle on the right side varied in steps of $\frac{1}{8}$ from standard, or $4\frac{2}{8}$, to $3\frac{3}{8}$ inches.

It proved unnecessary to include neutral *T*-objects which were larger than the standard. With some subjects it was equally unnecessary to present the smallest comparison objects. In a crude form we used the psychophysical procedure of constant stimuli; i. e., after satiation the various pairs of *T*-objects were shown in an irregular sequence, the pairs for which the direction of the judgments varied were given repeatedly, and from a table of the judgments the most probable point of equivalence was inferred. Actually this process required more than usual patience for both experimenters and subjects. We could not give the subjects merely one inspection period and then gather their judgments in a rapid succession of comparisons. Although after-effects as such are surprisingly persistent, their full amount is not preserved for long. Moreover, the fact should not be overlooked that repeated inspection of varying *T*-objects will also have after-effects which may influence the results in an unpredictable manner. For these reasons we used the following technique: an inspection period was followed by only a single comparison. After a rest period of at least fifteen minutes during which the subjects occupied themselves as they wished a second, equal, inspection period with a subsequent comparison was given, and so forth, until a sufficient number of judgments had been collected. Obviously, under these circumstances few subjects were able to complete their series in one day. But only with this procedure could the effect of the inspection period be in full force whenever two *T*-objects were compared. On the other hand, the rest period of fifteen minutes—or often much more—proved sufficient for the prevention of any accumulation or summation. In a prolonged series of comparisons no appreciable trend in the direction of the judgments could be observed. Each inspection period had the duration of forty-five seconds. Subjects observed from a distance of two yards.

Our measurements were crude because with our technique we had to restrict the number of judgments to the smallest number from which the equivalence point could be safely inferred. The following instance will give an impression of the degree of accuracy which was possible under these circumstances. One subject used—with reference to the object in the affected area—the categories *smaller* ($<$), *equal or slightly smaller* ($\leq$), *equal* ($=$), *equal or slightly larger* ($\geq$), and *larger* ($>$). The direction of his judgments varied only for three of the neutral circles which were compared with the affected constant object. For these three circles the distribution of his judgments was:

Diameters	$<$	$\leq$	$=$	$\geq$	$>$
$3\frac{5}{8}$	5			1	
$3\frac{3}{8}$	2	2	2	1	1
$3\frac{4}{8}$		1			5

It was concluded that the equivalence point lay between $3\frac{3}{8}$ and $3\frac{4}{8}$—to be sure, nearer the former value. Actually, 3.57 was taken to be the "true" value, which may be a trifle too low. With other subjects the procedure was the same.

Six subjects (A–F) took part in these measurements. For these observers the standard, or affected, object (D = 4.25″) was found to be equivalent to neutral circles with the diameters

Subjects:	A	B	C	D	E	F
	4.06	3.94	3.88	3.88	3.75	3.57.

In the affected area the diameter was therefore shortened by

.19	.31	.37	.37	.50	.68,

i. e., by $\frac{1}{5}$ to $\frac{2}{3}$ of an inch. In terms of area the loss was

9	14	17	17	22	29%.

These data confirm the reports of previous investigators who found, with other figural after-effects, that interindividual differences as to the amount of such effects are very great. They are so in the present instance. On the other hand, for every one of our subjects the size of the affected *T*-object is considerably reduced. Where the loss is smallest the shrinkage in terms of area is the same as though a sector of 32° had been cut from the *T*-circle. The corresponding sector for the strongest effect (subject F) would have an angle of 104°.

With reference to the specific purpose of this experiment we are now fully justified in concluding that a *T*-object which lies *within* an affected area loses much more of its size than does a *T*-object which coincides with the *I*-object. In the latter case we have never observed an effect that could even remotely compare with the results of the present experiment. In order to make sure that this holds also for the particular *I*-object which was used in these measurements we drew on a test chart two exact copies of this circle. Their arrangement with regard to the fixation mark was symmetrical, and the left circle had the position which the *I*-object had in the quantitative experiment. The *I*-object was again shown for forty-five seconds. After this period even the subject with the greatest effect in the quantitative experiment could not discover more than a just clearly perceptible shrinkage of the affected *T*-object.

It is, however, a simple differential experiment which gives us the most direct proof *ad oculos* that the size of a *T*-object suffers less in the case of coincidence than it does in the interior of an affected area. Figure 22 represents an *I*- and a *T*-pattern in the usual superposition. The ob-

long on the left side and the larger oblong on the right side are *I*-objects. The test chart contains two equal oblongs of which the left one coincides with the outline of the left *I*-object. Therefore, on the left side a figural after-effect under conditions of coincidence will be observed while on

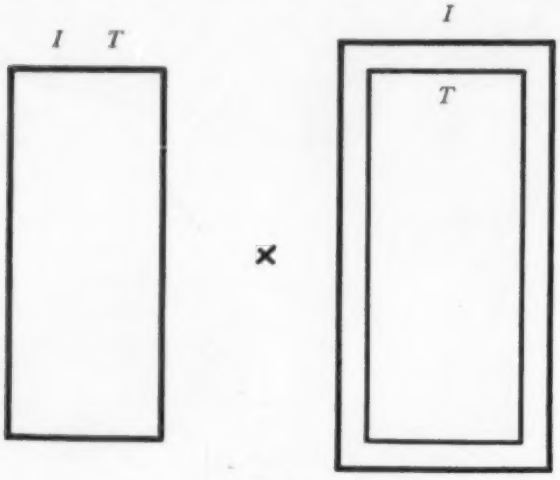


FIG. 22

the right side an identical *T*-object will be affected across a distance of one-quarter inch from the outline of the *I*-object.

The outcome of this experiment was perfectly clear. After a short inspection period the *T*-object on the right side began to look smaller, and as the inspection period was lengthened the difference gradually grew until at its final value it became very conspicuous indeed.

Undoubtedly, then, under the conditions of such experiments the size of *T*-objects is more strongly affected when the outline of the *I*-object surrounds that of the *T*-object than when both coincide. We are forced to conclude that the amount of this size effect does not simply depend upon the degree of satiation at the place of the *T*-object. If this conclusion were not accepted we should have to infer that with such *I*-figures satiation is greater at a distance from their outlines than it is at the place of these lines. This is not a plausible hypothesis. It is also contradicted by the behavior of other symptoms of satiation. All our observations indicate that the color of a *T*-object is more strongly affected when this object approaches the outline of the *I*-object. To our present knowledge the same holds for the depth effect, the receding of *T*-objects in the third dimension. In fact, even certain other size effects with which we are not now concerned follow this rule. Thus we have every reason to believe that satiation as such decreases as the distance from the outline of an *I*-figure increases. But on this ground our former assumption becomes unavoidable. In situations like the one of our quantitative experiment the size of a *T*-object does not depend merely upon the degree of satiation at the place of this object. Perhaps this fact *per se* is not surprising. After all, it follows from our own views that a *T*-object, like an *I*-figure, is functionally represented by a figure process which spreads beyond the area in which the color processes of the *T*-object are located. Hence, a *T*-object must be expected to react to "abnormal" conditions both in its environment and at its own place.

Several conclusions may be derived from this discussion:

First, it is now obvious why any comparison between a *T*-object and a thermometer must be misleading. A *T*-object is associated with an active process of its own whose distribution depends upon conditions in a region of some width. A good thermometer, on the other hand, passively indicates the temperature of a medium at the place at which the instrument is inserted.

Secondly, there is every reason to insist upon the distinction between figural after-effects and satiation, since a *T*-object that lies wholly in a region of maximal satiation need not show a strong after-effect in every respect.

Thirdly, certain symptoms of satiation actually *are* strongest in regions in which the highest degree of satiation has been established. It follows that the size effect which we have just measured in one instance cannot be explained in terms of those other symptoms. We shall readily grant that if an object's color or its localization in the third dimension are affected its size, too, may change merely in consequence of those facts. But this cannot be proposed as a general interpretation of the size effect which is now under discussion. For when a *T*-object is removed from the region of maximal satiation we find those other symptoms weakening while the size effect grows. Obviously, then, the various symptoms which are caused by satiation are independent of one another.

All this, of course, does not deprive our observations of their paradoxical nature. If we had merely to explain that at a distance from the places of highest satiation a *T*-object may still exhibit a strong size effect, we could indeed point

to the fact that a figure process satiates a considerable area, and that a subsequent figure process, that of a *T*-object, will also depend upon conditions in such a wide area. No more would be needed as an explanation. But the paradox with which we are confronted is much more serious than that. At a distance from the outline of the *I*-figure this size effect is actually much *greater* than it is when the *I*-object and the *T*-object coincide. For this finding no explanation has thus far been given. And yet the further we proceed the clearer will it become that precisely with this fact we have touched upon a basic principle of figural after-effects.

II

We shall next report about experiments which seemed to us at first quite unrelated to those of the first chapter. They were undertaken when the paradox which we have just been discussing had not yet been recognized. In fact, it was a mere chance observation which caused us to turn in what then appeared to be a new direction. Actually, the phenomena which we shall now describe could have been predicted if at the time that paradox had been fully appreciated.

For certain purposes we wanted to compare the size effects on the left and on the right side of Figure 23. But when the experiment was

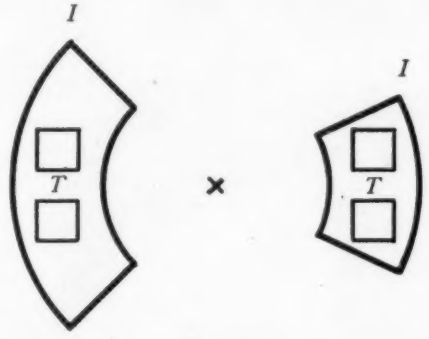


FIG. 23

performed our attention was first of all called to the fact that after the inspection period the squares on the right side were *distorted*. The upper horizontal edge of the square on top and the lower horizontal of the other square seemed to converge toward the right.

The connection between this fact and the argument of the preceding section may not at once be obvious. But the following experiments will make the relation quite clear.

It seemed advisable to subject our accidental observation to analysis under simplified conditions. Figure 24 in which the angle (45°) is the *I*-object, and the square the *T*-object, appeared to contain all characteristics of Figure 23 which could be responsible for the distortion. To most subjects Figure 24, which gives the *I*- and the

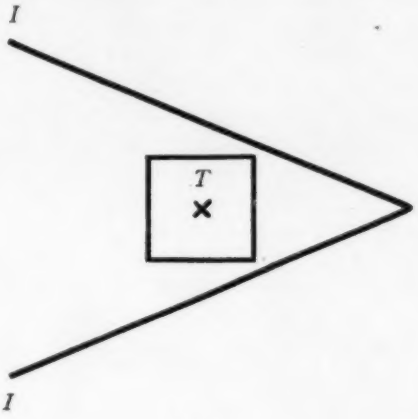


FIG. 24

T-object in one design, exhibits an immediate illusion inasmuch as the objectively horizontal edges of the square will be seen as slightly convergent toward the open side of the angle. The existence of such an immediate illusion in a similar pattern has been discussed by Ponzo. We were not acquainted with Ponzo's work when we chose this particular arrangement.[9]

In our experiment the angle alone was of course shown first, and the *T*-object only when the chart with the angle had been removed. We found that after very short inspection of the angle the square was clearly distorted in the expected sense; it appeared as a trapezoid with shorter vertical facing the open side of the angle. We had the impression that this distortion went considerably farther than it does when the *T*- and the *I*-object are given simultaneously (Ponzo's illusion). If both the *I*-chart and the *T*-chart are turned by 180° the same effect will be observed, i.e., the *T*-square converges toward the open side of the angle, irrespective of the orientation of the angle in space.

In view of our recent argument this observation can be interpreted as one more instance of the paradox with which we are already ac-

[9] A recent investigation by W. R. Sickles which develops a particular theory of Ponzo's illusion called our attention to Ponzo's work. *Cf.* Sickles (23) and Ponzo (21). Sickles' theory will be discussed in another chapter.

quainted. On the right side the horizontal edges of the *T*-square lie nearer the sides of the *I*-angle than they do on the left side. In one detail, it is true, this situation differs from the arrangement which first revealed the paradox. We then found that the size effect is smaller when the *T*-object and the *I*-object *coincide* than it is when the *T*-object lies within the area of the *I*-object. But it seems natural to assume that the paradox is not restricted to this particular condition; in other words, that within certain limits the size effect decreases as the outlines of the *T*-object *approach* those of the *I*-object. If this generalization be granted our observation follows immediately: The vertical distance between the *T*-square and the sides of the *I*-angle grows from the right to the left. Consquently in the vertical direction the square must shrink less on the right side than it does on the left side, and under these circumstances its horizontal edges will necessarily converge toward the left. We do not mean to imply that at each point from the right to the left the vertical size of the square is determined merely by the distance which separates the two objects at that particular point. Surely, a given figural after-effect does not consist of independent partial effects which single points of an *I*-figure exert upon single points of a *T*-object. Not only theoretical considerations but also numerous experimental facts make any such assumption entirely impossible. In a molar sense it remains nevertheless true that on the right side the edges of the square are nearer the sides of the angle than they are on the left side. And, again in a molar sense, this condition varies in a continuous fashion. No wonder then that as we proceed from the right to the left the vertical size of the square is progressively reduced.

So long as no further evidence is added our interpretation will perhaps seem slightly artificial. But we are confident that this impression will soon disappear. To be sure, a different explanation would be given by Gibson. He would say that our observation only confirms his discovery of the "tilted line" effect. Indeed, the sides of our *I*-angle are straight lines which form acute angles of moderate size with the horizontal direction. According to Gibson actually horizontal *T*-lines will turn in space when such tilted lines have previously been inspected. Undoubtedly, he would apply this rule to the horizontal edges of our *T*-square; the direction in which these edges are found to turn in our experiment is just the direction in which *T*-lines turn in his effect.

Now we are quite convinced that in this experiment a first contact between Gibson's discoveries and our own observations is actually established. Also, no objection can be raised against the factual basis from which Gibson's concept of a "tilted line" effect is derived. However, although this concept refers to definite facts it also contains an interpretation of such facts; and in another chapter it will be shown that this interpretation is not justified. Under these circumstances the discussion of Gibson's experiments and principles must be postponed even if a few of our next observations are clearly related to his discoveries.

Without any special discussion it will now be recognized that our first, accidental, observation (Fig. 23) allows of the same interpretation as has just been given to the present experiment. That first observation shows, however, that it makes no great difference whether a closed or a partially open outline is given as an *I*-figure. This deserves mention because otherwise one might suspect that the argument of section I does not really apply to an open angle, and that therefore our interpretation is arbitrary. Actually in Figure 24 itself the angle may be closed by a circular arc or by a straight vertical line. In both cases a *T*-square within the sector or the triangle will still be found to converge toward the wider side of the *I*-figure.

On the other hand, the effect is not restricted to the particular angle of 45°. We obtained similar results when the angle was reduced to 30° or increased to 90° (Fig. 25). In the latter case the convergence of the objective square toward the right side was quite conspicuous although with this wider angle the after-effect seemed to subside more quickly.[10]

When the outline square was replaced by a solid square the after-effect did not disappear. It may be that the horizontal edges now converged less than did those of an outline square. Also, with such a *T*-object the effect seemed concentrated in parts of the square which were not

[10] It will be observed that here and in the next Figures Ponzo's immediate illusion is again noticeable. The fact that in combinations of *I*-figures and *T*-objects the latter often show immediate distortions which are analogous to corresponding after-effects will be discussed in Chapter Seven. In larger drawings these immediate distortions are obvious. They are less striking in smaller reproductions such as the present Figures.

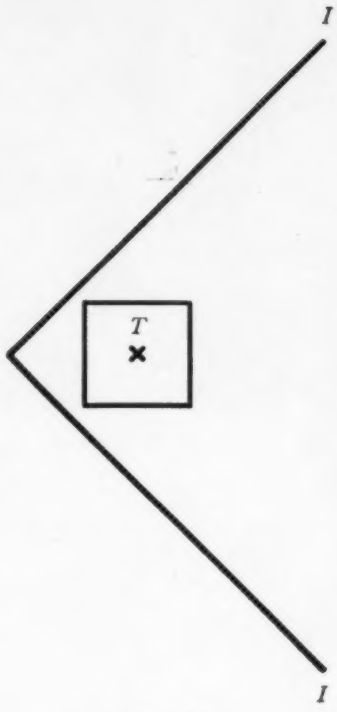


FIG. 25

too distant from the sides of the angle. As a consequence the edges of the *T*-object appeared slightly curved rather than simply turned in space, the curvature being concave towards the sides of the angle. This phenomenon also appeared in Figure 26, in which the left side of an

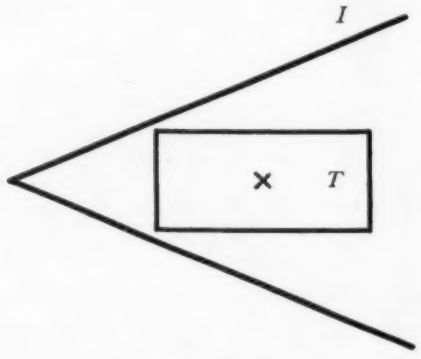


FIG. 26

outline oblong almost touches the angle. On the left side the horizontal lines were clearly curved in the same sense. From the point of view of our interpretation there is of course no reason why the shrinking of the *T*-object should under all circumstances be a linear function of the growing distance from the *I*-lines.

With the following variations of the *T*-object we turn to experiments which are more directly related to our interpretation of the present facts. On the face of it the situation investigated in section I of this chapter differed considerably from the situations with which we are now concerned. And yet, in interpreting the latter we have used a principle that was taken from the former. We shall now proceed to observations which will serve as intermediate and connecting steps. Thus the relevance of the size effect in section I to the distortion of rectangles and squares within an angle will become more plausible.

In the first place, we now use two *T*-objects instead of merely one. With this variation we approach the situation which has been discussed in section I: In Figure 27 the left oblong, as a

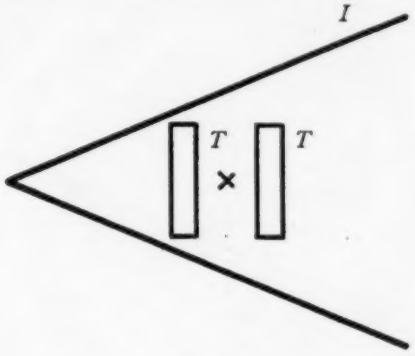


FIG. 27

single object, extends into the immediate neighborhood of the *I*-lines; the right oblong, another separate object, remains farther apart from these lines. At the same time the two oblongs may be considered as remnants of a large square the horizontal edges of which would surely converge toward the open side of the angle. This is the fact to which our interpretation refers. If this interpretation is valid the two oblongs must have different heights after the angle has been inspected. Actually, this is just what the subjects reported. The left oblong was higher than the right one. It will not be denied that this observation may be regarded either as an instance of the paradoxical size effect or as a particular form of the effect with which we are occupied in this section.

The subjects did not say whether in each oblong the short horizontal edges were slightly convergent toward the right. These edges may have been too short for such an effect to be

easily noticed. But about another question there was no doubt whatsoever: The oblong on the left side appeared much paler than the one on the right side; it also lay farther back in space. This demonstrates the independence of the various symptoms of satiation to which we referred in section I (p. 284). Near the apex of the angle satiation must be greater than it is near the opening of the angle. Accordingly both the *color*- and the *depth*-symptom are present in the former region while they are absent or much less pronounced near the open side of the *I*-object. None the less the oblong which is nearer the apex shrinks less than does its partner which lies in a more weakly satiated region.

In order to avoid monotonous repetition the independence of the symptoms will as a rule not be mentioned in further reports. It was regularly observed when our patterns allowed it to appear.

When the oblongs are reduced to two verticals (Fig. 28), a configuration which resembles Ponzo's illusion more closely, inspection of the

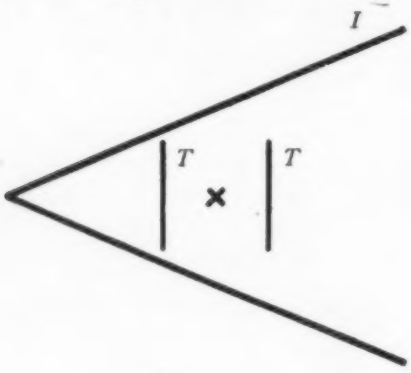


FIG. 28

angle has the same effect: the left vertical is longer than the right one. Once more we insist that this observation refers just as much to the facts which we are now investigating as it refers to the size effect and its paradoxical behavior (section I). Any lines which we draw or imagine as connections of the ends of the verticals will of course converge toward the opening of the angle.

An interesting modification of these tests is presented in Figure 29 in which the *T*-pattern is the familiar square of small squares. In this instance observation shows that distances between *T*-objects follow the same paradoxical rule as we know to hold for closed outline objects. After the inspection period the squares on the left side are clearly more distant from one another than are the squares on the right side. It may be remarked that with four solid squares in the same *T*-pattern the outcome of the

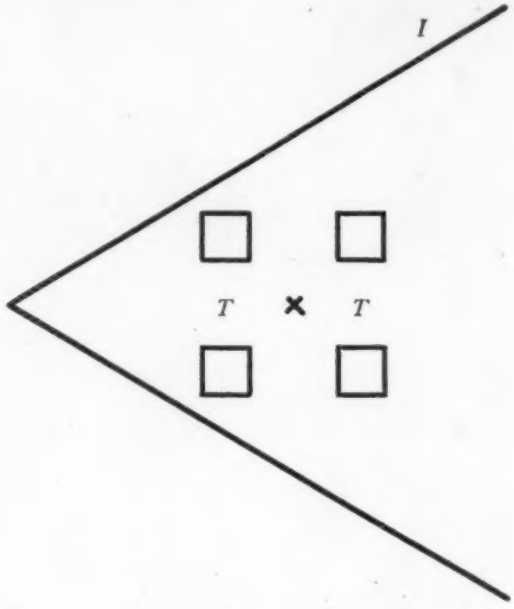


FIG. 29

experiment is not essentially altered. In some observations with this arrangement the horizontal edges of the small squares, at least those which face the sides of the angle, are actually found to turn in the expected direction, i. e., in the sense of the convergence of the whole *T*-pattern.

Incidentally, in these experiments figural after-effects show again the curious asymmetry in the vertical dimension which was first mentioned on page 281. The lower horizontal lines tend to turn more strongly into an oblique position than do the horizontal lines above the fixation mark. With two separate *T*-objects the difference as to vertical height is also localized mainly at their lower ends, one of which appears higher than the other.

As a second step of this analysis we shall now clarify a point at which our description of the facts may seem to have gone farther than they warrant. "The *T*-object looks comparatively small in regions in which the previously inspected *I*-object had considerable width, and the *T*-object appears comparatively big in places in which the *I*-figure was narrow." Would this not suffice as a description of our observations? Moreover, such a description would immediately suggest to us the term "successive contrast."

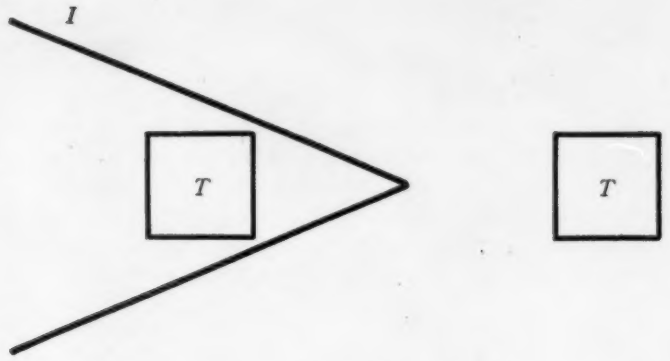


FIG. 30

Why should we worry about an alleged paradox if this familiar term covers our facts?

The trouble with such statements is that they prefer an atmosphere of comfortable vagueness to a clear light in which problems appear sharply outlined. If we were to accept the present proposal, what exactly would follow for the size of *T*-objects within an angle? Granting that the relation between sizes near the apex and sizes near the opening of the angle is indicated, what are these sizes when compared with those of objects on a neutral homogeneous ground? From that description of the facts one might, for instance, infer that near the apex of the angle an object is taller than such a standard while near the opening it is shorter than the same yardstick. If, on the other hand, the statement is not meant to have this implication, what is its actual import with reference to the size of objects which do not lie within an angle at all?

We prefer to return to our own interpretation which makes in this respect a definite assumption. If the interpretation is right, objects within the angle will always be shortened whether they lie near the open side or near the apex; and their sizes will differ merely because this effect is, paradoxically, greater near the open side. This conclusion can be tested with the arrangement of Figure 30 in which a square within the angle is compared with an objectively equal square on neutral ground. The experiment gives a clear decision. The square within the angle does not only converge toward the open side of the angle; it is also, including the vertical edge near the apex, distinctly smaller throughout. Hence we have no reason to correct our description of the facts; it implies no more than is justified by our observations.

As a final step we shall now eliminate from our situation any obliqueness in the spatial relation of the *I*- and the *T*-pattern. At the same time we shall divide the *I*-figure into two separate parts. It will be difficult to decide whether in the arrangement of Figure 31 the *I*-pattern is still to be regarded as a modification of the angle or rather as a modification of the situation which we considered in section I. As we see it the present pattern falls under *both* categories.

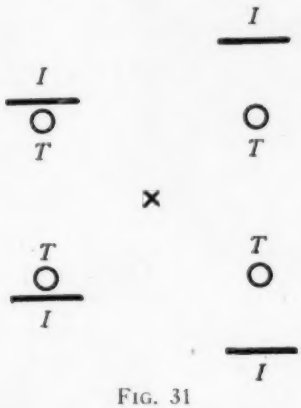


FIG. 31

Accordingly, experimental results will here refer to a situation which still contains an essential characteristic of the angular pattern, while at the same time oblique lines are absent and an interpretation in terms of section I is obviously indicated.

After the inspection period the two *T*-objects on the left side are clearly farther apart than are the two on the right side. The paradoxical nature of the size effect could hardly be better demonstrated than it is in this observation. One cannot raise the objection that, taken as a whole, this *I*-pattern still contains an element of obliqueness inasmuch as it may be seen as converging from the right to the left. We do not

deny that this is possible or, perhaps, natural. In fact, for just this reason the pattern remains a modification of the angular pattern. But it is quite another question whether that obliqueness as such, without any actual oblique contours, causes the present differential effect. For this to happen, a figural after-effect in the left half of the field would have to depend strongly upon the position of the inspection objects in the right half, and vice versa. In our experience there has never been an indication of such an interrelation of after-effects across the median vertical when the horizontal distance was as great as it is in this experiment.

If only the left half of the *I*-pattern is shown for inspection we can once more test whether vertical distances are actually shortened on both sides, as our interpretation contends. This is indeed the case. The small circles on the left now appear nearer each other than do their unaffected partners on the right side. Thus the distance which, paradoxically, is greater in the differential experiment proves none the less to be shorter than an unaffected standard.

It will almost be unnecessary to remark that within certain limits the outcome of such experiments does not depend upon the particular nature of the *T*-objects. The four circles may, for instance, be replaced by two vertical lines (Fig. 32). After the inspection period the vertical on the left side is longer.

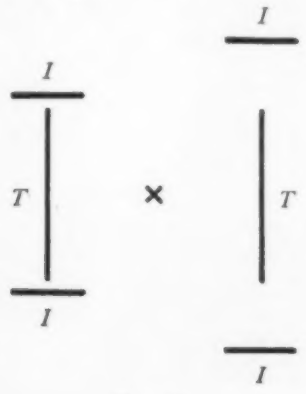


FIG. 32

We trust that after this analysis our interpretation of after-effects within an angle will no longer appear artificial. Owing to its paradoxical nature the size effect must cause just such phenomena within an angle as are actually observed. So far, then, no new principle seems to be involved in these after-effects. Conversely, of course, the paradox is once more verified by the experiments of this section.

CHAPTER THREE

DISPLACEMENTS

I

Our analysis has proved that all figural after-effects which have been described in Chapter Two follow from only one principle. In the present chapter we propose to show that our formulation of this principle can be greatly improved; in fact, that it must be improved because in its present version it stresses one particular phase of our observations, ignores others, and thus causes a biased view of the situation. The over-emphasized item is the *size* of *T*-objects. In our very first experiment (p. 271) we found, besides such symptoms of satiation as the color- and the depth-effect, an unexpected third symptom: the shrinking of the *T*-object. Had the first experiment been done with a different pattern our attention might have been focused upon another spatial factor, a factor that always accompanies a loss of size, but has so far barely been mentioned in our reports. This factor is *change of location* or *displacement*.

From a phenomenological point of view, it seems, the size of an object in a given plane cannot change unless at least parts of its boundary change their location at the same time. This applies to instances in which the size of a *T*-figure is affected by previous satiation just as it does to a decrease which is caused by a change of retinal stimuli. As a consequence all our experiments in which *T*-objects were found to shrink were at the same time experiments in which outlines or contours shifted from one position to another. Consider the measurements reported in section I of Chapter Two. We found that under the conditions of those measurements the diameter of a *T*-circle, which lay within the area of an *I*-circle, was shortened by about one-fifth of an inch with one subject and by much more with others. Is it too bold an inference if we conclude that all around the *T*-circle the location of its outline was shifted by one-half of that amount, i. e., by one-tenth of an inch and, with most subjects, by much more?

At first it may appear to be a mere matter of form whether we describe such facts in terms of change of size or in terms of displacements. However, a very simple consideration will show that actually this distinction has considerable factual relevance.

In the experiments of Chapter Two the *T*-figures were either enclosed in the outline of the *I*-object, or they lay between two parts of an *I*-pattern. Under such conditions, we stated, the *T*-figures invariably shrank. When expressing this effect in terms of displacement of outlines we should have to say that *the outlines of those figures receded from regions in which a high degree of satiation had been established*. For the moment we need not lay much stress on the causal implications which this new formulation may suggest. But we cannot overlook the fact that from this point of view the "size effect," in the sense of a decrease of size, may prove to be a special form of a more fundamental phenomenon. For, if the new formulation is correct, *T*-objects may not shrink under all circumstances. Rather, in certain configurations a *T*-object may be simply *displaced* while its size is not, or not essentially, affected. In other patterns a *T*-object may not shrink but, on the contrary, *grow* when its outline recedes as a whole from a zone of maximal satiation. And in the observations of Chapter Two every *T*-object may have *shrunk* merely because we were using particular patterns in which the receding of outlines could have no other effect.

Indeed, if the *T*-object is surrounded by a zone of intense satiation, or if it lies between two such zones, displacement of its outline away from those affected regions will necessarily make the object shrink; because in receding from those regions the various parts of the outline must approach each other. Under these circumstances it may indeed appear as though the after-effect could be described as we choose, either in terms of shrinkage or in terms of displacement. Suppose, however, that the *T*-object lies on one side of the *I*-object while there is little or no satiation on the other side of the *T*-figure. Such a situation, it will be realized, may lead to after-effects which are not covered by our term size effect. For instance, if in this situation *all* parts of the outline of the *T*-object should recede from the affected area it would no longer suffice to mention any shrinkage that may or may not accompany the displacement of the object. For, quite apart from such a size effect, the *T*-object would also be shifted as a whole.

In view of this possibility it is perhaps not advisable to concentrate too much upon displacements which the outline of a *T*-object may suffer; because satiation on one side may displace the *T*-object as a body, its outline being of course included in this process. In fact, we should prefer a description in such terms if a *solid* figure were used as a *T*-object, and if that solid figure, too, withdrew from an affected region. Obviously, in such a case it would be unnatural to emphasize the displacement of the contour as though the displacement in terms of area were only a secondary effect.

So far we have been discussing mere possibilities. It remains conceivable that displacements of outlines and other parts of *T*-objects are restricted to instances in which these objects shrink. However, the following evidence decides against such a view. When a *T*-object lies on one side of an affected area the object does recede from that region.

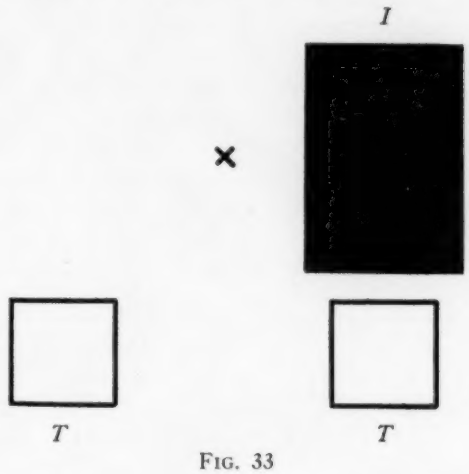


FIG. 33

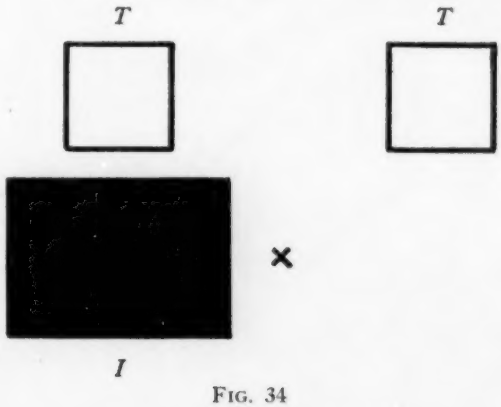


FIG. 34

After inspection of the oblong in Figure 33 the two *T*-squares will not appear at the same height. The one on the right side will not only be pale; it will also lie lower than the other. In Figure 34, on the other hand, the square on the left

side will appear higher. Attention should of course be given to that outline of the *T*-object which does not face the satiated region. It will be found that, whether or not the size of the object is at all affected, in these experiments *both* its horizontal edges are displaced in the same direction. As a matter of fact, precisely the displacement of the more distant edge of the *T*-object is sometimes the most obvious part of the observation.

An arrangement just like the ones with which we are now dealing was used once before, although for another purpose. In Chapter One (p. 275) an attempt was made to determine how far away from the area of an *I*-object symptoms of satiation are still observable. Two squares were then used as *T*-objects, and the *I*-object was an oblong that lay above one of these squares. In those experiments the subjects sometimes mentioned that, quite apart from color-, depth-, and size-effects, there was a fourth symptom: the affected square was displaced as a whole. The direction of this effect was the one which has just been indicated.

Further proof that such displacements are independent of any shrinking of the *T*-objects will be found in experiments in which a mere line is the *T*-object. Such a line will often be displaced by amounts which are large in comparison with its own width (*cf.* below p. 304).

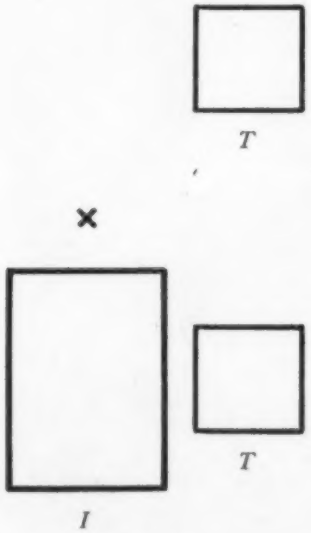


FIG. 35

Displacements occur not only in the vertical dimension. After the inspection period the *T*-objects of Figure 35 appear in an oblique position: the lower one lies too far to the right side. This observation also proves that a *T*-object recedes from a satiated region whether the satiating figure has been a solid or an outline object. Conversely, if a solid figure is used as a *T*-object it will show the same displacement away from an affected area.

The displacement of *T*-objects can be demonstrated in a more conspicuous fashion. So far the position of an affected *T*-square has been compared with that of an equal *neutral* square which objectively either lies just underneath or has the same height. In the following arrangement (Fig. 36) both *T*-squares are affected; but

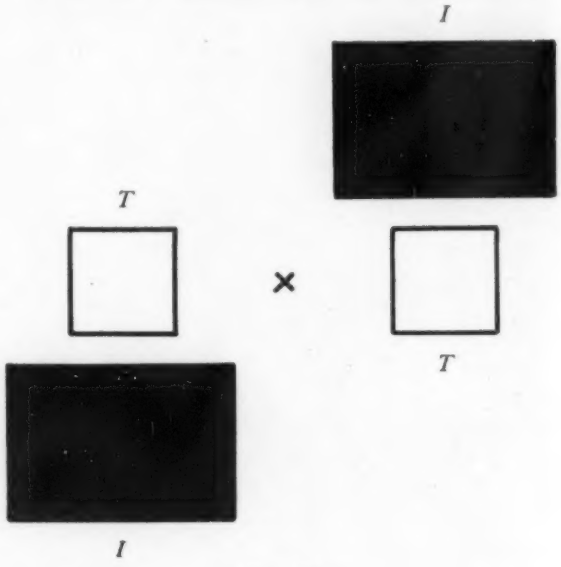


FIG. 36

in receding from the affected areas in their neighborhood they move in opposite directions. Thus their oblique position becomes more striking than that of the *T*-squares in our previous experiments.

In the next section we intend to show that the displacement of *T*-objects in the neighborhood of an affected region can be measured. But the qualitative observations which have just been described will be sufficient as evidence that our suspicion was justified: *T-objects recede from satiated areas.* This effect persists when conditions exclude any description in terms of loss of size. In the next pages this new principle will be used both for the deduction of further facts and for an interpretation of former observations.

If a *T*-object is properly placed *within* the

area of a previously inspected *I*-figure it will lie between the strongly affected zone near the outline of this figure and its less satiated interior. We may expect that under these circumstances, too, the *T*-object will recede from the zone which is more strongly satiated. This is actually the case. In Figure 37 the *T*-object within the

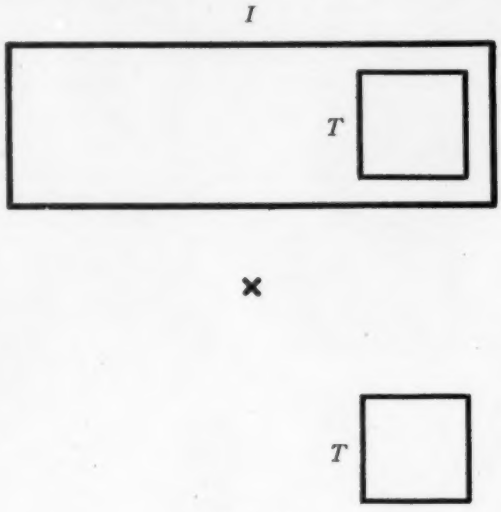


FIG. 37

area of the *I*-oblong is not seen straight above its neutral partner. Rather it appears shifted to the left. One cannot object that an interpretation in terms of shrinkage is applicable to this fact inasmuch as the distance between the *T*-object and the left edge of the oblong is shortened. When the *T*-objects are shown that edge of the *I*-oblong is no longer visible, and therefore the observed position of the affected *T*-object is not referred to that edge. Moreover, the proposed interpretation is arbitrary. For, if the distance between the affected *T*-object and the left edge of the (now absent) oblong is decreased, the distance between that object and the right edge of the oblong will be just as much increased.

To be sure, some reference to factors besides the affected *T*-object itself will always be involved when we state that this object is displaced. For instance, in describing the last few observations, we referred not only to the second *T*-object but also to the main directions of space, the horizontal and the vertical, of which we are normally aware in any visual field. Our pairs of *T*-objects appeared ***oblique***. But, generally speaking, a given displacement may be revealed by changes in one kind of reference or another. It depends merely upon our choice of the *T*-patterns which of these possibilities will actually be realized. A displacement to the left, for instance, showed as obliqueness of a pair of *T*-objects when the arrangement of Figure 37 was used. If, in the direction of the displacement and at a great distance, a third *T*-object had been shown, and just underneath it a fourth (*cf.* Fig. 38), then we should probably have found not only that the right pair of *T*-objects was in an oblique position but also that the objects above the fixation mark lay nearer each other than the other two. This size effect would have been greatly enhanced if the *I*-oblong had been extended to the left so as to include both upper squares symmetrically. For, under such circumstances these objects would have been displaced toward each other as each would have receded from the nearest edge of the oblong (*cf.* Fig. 16). In this fashion the concrete references the alterations of which we call figural after-effects may be varied. But while they vary, one fact persists as an integral phase of *all* distortions. This fact is displacement. Therefore, if we wish to stress the invariable, the gen-

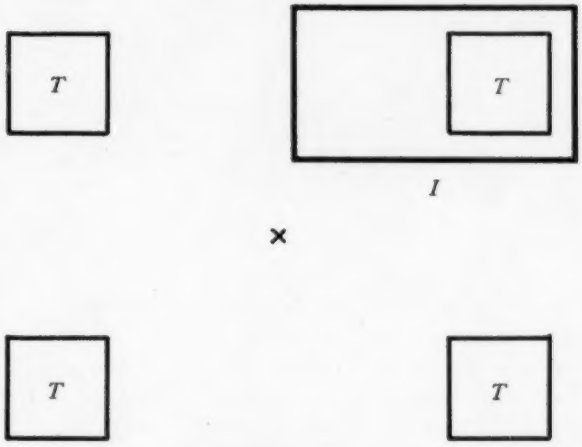


FIG. 38

eral, and in this sense the most fundamental, principle of such after-effects we shall have to emphasize that *T*-objects or parts of these objects are displaced. Both losses of size and obliqueness of *T*-objects are of course actual facts in certain concrete after-effects. But they are merely particular facts which depend upon particular situations, while every spatial symptom of satiation involves displacement in space.

We hope that this principle will not be wrongly interpreted. We do say that, as a ***genus***, dis-

placements are involved in all distortions of *T*-patterns. But we cannot formulate it as a general rule that in a satiated field a given object will suffer a certain amount of displacement irrespective of the particular total *T*-pattern in which that object is shown. Occasionally observation may agree with such a rule (*cf.* Fig. 40). We know, however, that the location of objects within a larger pattern is generally influenced by this pattern. Patterns often do something to their parts. That much follows from numerous "illusions." Correspondingly, a given pattern may facilitate certain displacements which would hardly show in other configurations. Thus, displacements in satiated fields are surely not invariants in the *quantitative* sense of this term.

We are now prepared to introduce a further phenomenon. In certain arrangements satiation *enlarges* the distance between *T*-objects. If the principle of displacement were not yet recognized this phenomenon would appear as a special figural after-effect. Actually it is only one more consequence of the fact that *T*-objects recede from satiated areas.

From the point of view to which we adhered in the first two chapters no such effects could be predicted. Suppose we were still thinking in terms of the size effect. This effect, we should then have to say, consists in a shrinking of *T*-objects or of distances between such objects; and it occurs when these objects lie in sufficiently satiated areas. In fact, this was the sense in which the "size symptom" was conceived when a short report about our first experiments was elsewhere given (17, ch. II). Now, *T*-objects which lie within a previously inspected figure fulfil that condition; they are placed in a satiated area. But we remember, that satiation spreads *beyond* the outline of a figure into its environment. If therefore two *T*-objects lie on both sides of an *I*-figure, but not too far away, they are also placed within an affected region, and the same holds for the distance between them. Under such circumstances this distance may be expected to *shrink*. It would simply appear to us as one more instance of the familiar size effect if observation were to confirm that expectation.

From our new principle, however, just the opposite effect must be predicted. According to this principle *T*-objects within the area of an *I*-figure approach each other inasmuch as they recede from zones of maximal satiation, the zones about the edges of that figure. Obviously, then, if a zone of intense satiation lies *between* two *T*-objects the same basic phenomenon, the receding of these objects from zones of intense satiation, must lead to an *increase* of their distance.

That this is the correct prediction can be shown in very simple experiments. In Figure 39, for instance, the *T*-circles on the left side lie

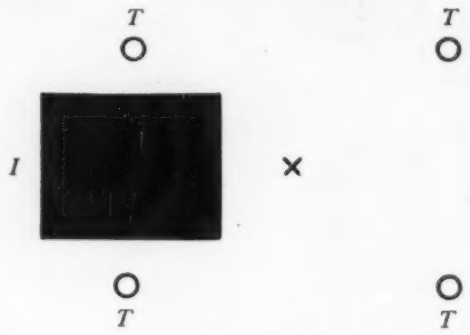


FIG. 39

farther apart than those on the right side when the *I*-square has previously been inspected. It deserves mention that in this observation only the size effect reverses its sign while other symptoms of satiation preserve their usual character: the *T*-objects on the left side lie farther back in space and are paler than their partners.

The effect persists when the *T*-circles are replaced by *T*-squares. It also persists when the distance between the *T*-objects and the *I*-figure is varied within certain limits. Moreover, if satiation between two *T*-objects is caused by

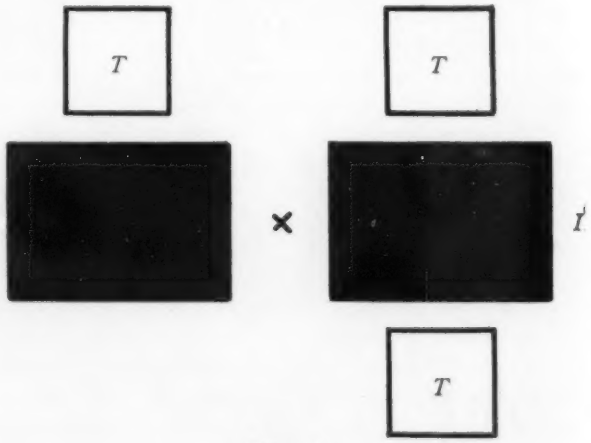


FIG. 40

an outline *I*-figure their distance will again be found to grow.

It may be asked whether other factors besides the displacement of each object taken separately contribute to the increase of their distance. For instance, when two *T*-objects are given, satiation of the area between them might also influence their distance as such, quite apart from the fact that the objects are displaced. The answer seems to be that, at least under the conditions of the present experiments, no such additional factor is involved. We do not find that the upper *T*-square on the right side of Figure 40 lies higher than the corresponding square on the left side. Presumably it would lie higher if a special increase of the distance were added to the displacement of the right *T*-objects.[11]

When the vertical distance between two *T*-objects shrinks—as it does with *T*-objects within an *I*-figure—this size effect does not affect both objects to the same extent. It will be remembered that under these circumstances the lower object tends to be more strongly lifted than the other is lowered (*cf.* p. 281). In the present observations, too, visual space proves to be anisotropic. Again it is the lower *T*-object which as a rule appears more strongly affected. Just as in those earlier experiments it receded further upward from the next edge of the *I*-figure than the higher *T*-object receded downward, so the displacement downward of the lower *T*-object is now greater than the displacement upward of its partner.

In view of these data we realize once more that no formulation in this field which stresses any particular size effects can be regarded as satisfactory. Only a formulation in terms of displacement applies quite generally.

As to size effects, our very first observations might have taught us that under certain conditions satiation must *increase* given distances. For, if the size of an object decreases, its boundary moves in a certain direction, namely, toward the center of the object. It is therefore fairly obvious that if we introduce a neutral landmark farther off that center, the distance between the boundary and such a landmark must grow as the boundary moves inward. In this sense the very shrinking of an object implies that distances between this object and certain other things increase.

The fact that satiation may lead both to an increase and to a shrinking of given distances allows us to construct a demonstration experiment which some observers find particularly impressive. Instead of comparing an unaffected distance either with an increased or a decreased one we may, in one observation, contrast the reduction of a given distance on one side of the field with the growth of that distance on the other side. In Figure 41 the two *T*-objects

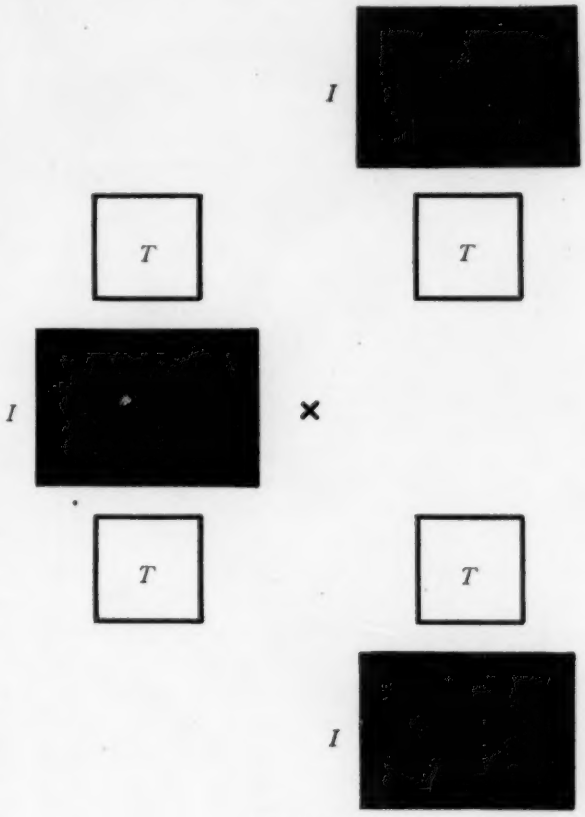


FIG. 41

on the right lie between two strongly satiated areas. Therefore they recede toward each other. On the left side, however, a satiated area lies between the two *T*-objects. Consequently in receding from this area these objects also recede from each other. The two changes which occur in opposite directions combine to cause a fairly striking effect. (We recommend that in a repetition of this experiment the dimensions of the pattern, particularly the distance between the *T*-objects and the *I*-figures, be chosen in agreement with footnote 3, p. 271.)

At the risk of formulating expressly what may at this point appear fairly obvious we propose briefly to return to the first chapters of this report. All size effects which were described in those chapters had one point in common: sizes or distances were found to shrink when the

[11] On page 294 reasons have been given which ought to prevent any rash generalization of the present finding.

medium was satiated. Our present analysis applies directly to a certain number of these instances, first of all to those in which two *T*-objects lay within, but rather near, the outline of a closed *I*-figure while the region between them was less affected (*cf.* Figs. 14, 15, 16). But the situation remained essentially the same if the *T*-objects lay between two *separate* *I*-objects which were a sufficient distance apart. In this case, too, the region between the *T*-objects was less affected than their environment in the outside direction, and as a consequence when they receded they had to approach each other (*cf.* Figs. 29 and 31). The paradoxical fact that the distances between *T*-objects shrank more when these objects lay farther apart from the *I*-outlines will occupy us later. At any rate, the *direction* of their displacement and therefore also the sign of the resulting size effect agreed with the principle of displacement which we have just established.

But within the area of an *I*-pattern a *closed* *T*-figure behaves exactly like a pair of such figures. Just as the distance between two *T*-objects is decreased so any analogous distance within a closed *T*-figure becomes shorter when the size of this figure shrinks. Clearly, mere consistency demands that under these circumstances the same interpretation be given in both cases. We are, in fact, convinced that in such situations closed *T*-objects shrink mainly inasmuch as their boundaries recede from surrounding zones of intense satiation. However, with a closed *T*-figure, too, a certain condition must be fulfilled if our interpretation is to be applicable. The outline of the *T*-object must lie within the area of the *I*-figure in such a way as to make the interior of the *T*-object a place of less satiation than obtains in the environment of this object. Among the experiments described in the first two chapters a further large group is thus included in our interpretation (*cf.* Figs. 5, 6, 11–13, 21, 22, 24–27, 30). We again ignore the fact that the shrinking of closed *T*-objects exhibits the paradox which has been discussed in Chapter Two. This paradox refers to the amount, not to the direction, of the effects in question. And it is only with their direction that we are now concerned.

It will be remarked that in our present survey we do not include *all* experiments in which the size of a *T*-figure—or the distance between two such figures—was found to shrink. In a certain group of our observations the asymmetrical situation to which we have just referred did not obtain. Rather, in these situations the *T*-objects lay in a homogeneously satiated region. And yet the *T*-objects—or the distances between them—were found to shrink. An interpretation of this fact will be given later.

To complete our survey of the facts now under discussion we must, however, perform one more experiment. Closed *T*-figures within the area of *I*-figures shrink mainly because their boundary and adjacent parts recede from strongly satiated regions in their neighborhood. In this section we have shown that when a region of intense satiation lies between two *T*-objects these objects will also recede from that zone, and that their distance will *grow* accordingly. It remains to be proved that a *closed* *T*-figure which completely *surrounds* a satiated area is also enlarged. We must indeed expect that, all around, the outline of this figure will recede

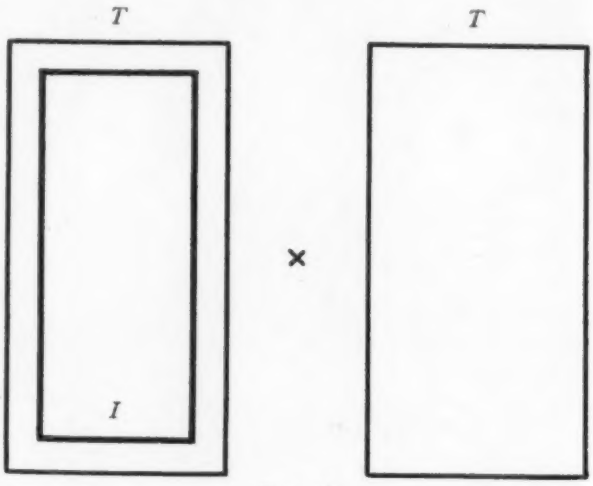


FIG. 42

from that area, and that as a consequence the size of the figure will grow. Our prediction can be verified with arrangements such as that of Figure 42. When after the inspection period the *T*-object on the left side is compared with the objectively equal figure on the right side the former will appear larger. A similar observation will be made if circles rather than oblongs are used. We wish to mention, however, that in our experience this particular effect has as a rule been short-lived. After the inspection period the *T*-pattern should therefore be shown without any delay.

We do not wish to burden our report with a discussion of details. The following observations, however, deserve to be mentioned. If the *I*-object is an oblong, and if two *T*-squares are shown within the affected area the distance

between these squares is shortened. On the other hand, if with the same *I*-object the *T*-objects are shown at a small distance, one above and the other below the oblong, their distance will be greater than the physically equal distance of two *T*-objects in an unaffected zone. Between these two situations the transition is not abrupt. Suppose that the *T*-squares lie half inside and half outside the area of the *I*-oblong. If the inspection period is not too short the distance between these squares will decrease. Even with *T*-squares which lie entirely outside the oblong but just adjacent to it the same will be observed—always provided that the inspection period has been sufficiently long (*cf.* 17, p. 99). In precisely the same situation, however, the distance between the *T*-squares may be increased if the inspection period is quite short. In other words, in this instance the direction of the effect depends upon the degree in which the medium has been altered by previous inspection. Also, during late stages of the inspection period a figure process may not affect the medium in quite the same way as it does in the beginning. Incidentally, in the same arrangement slight variations of the *I*-object may turn the effect in one direction or the other. It is true that with an oblong of small width the effect tends to lie in the direction of decreased distance between the *T*-objects when inspection has been sufficiently long. But with an *I-square* of considerable size it will be more difficult to obtain the same result unless the inspection period is very long indeed.

II

In formulating the principle of displacement we have practically ignored the paradox which was established in Chapter Two. We shall now return to this curious fact and relate it to the new principle.

The paradox consisted in this: within the area of an *I*-figure—or between two such figures—a *T*-object loses more of its size when its outline lies farther away from the region of maximal satiation. Under the conditions to which this formulation refers it is of course no more than a change of expression if we now say: within the area of an *I*-figure—or between two such figures—the outline of a *T*-object recedes more strongly from regions of maximal satiation when it is farther away from these regions. In the meantime, however, we have decided to regard displacements as fundamental symptoms of satiation, and to give a less central position to any particular forms, such as size effects, which displacements may assume in special configurations. From this point of view a displacement which goes with the shrinking of an object is not essentially different from any other instance in which a *T*-object recedes from a satiated area. *We must therefore conclude that the distance paradox is a general characteristic of all such displacements.* Thus it will be our next task to demonstrate the paradox, first, when *no* affection of size accompanies the displacement of a *T*-object and, secondly, when displacements *enlarge* given sizes or distances.

In Figure 43 the *T*-circle on the right side touches the edge of the *I*-object while between the left circle and the second *I*-object there is a

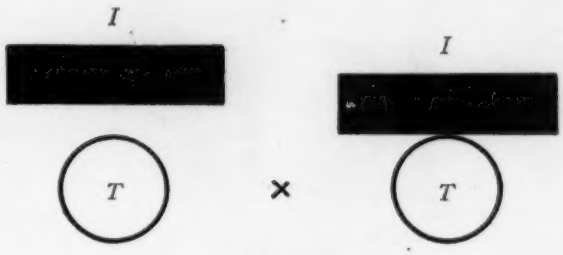


FIG. 43

distance of one-quarter of an inch. After the inspection period it will be found that the left circle appears lower in space than the right circle. Figure 44 demonstrates the same fact under conditions of horizontal displacement. When satiation has been established the lower *T*-object does not lie vertically below its partner; it appears shifted to the left. If two such paradoxical effects occur on both sides of the *I*-pattern the result will of course be more striking. At the same time the second point of our program will be settled in such an arrangement. For, under these circumstances the displacement of the *T*-objects will make their distance *grow*. In Figure 45 the *T*-circles on the right side appear farther apart than their partners on the left side although the circles on the left side lie nearer the *I*-object. This is a differential effect. If merely the left *I*-object is given, the

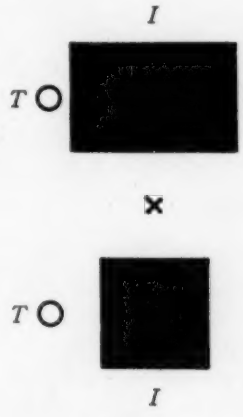


FIG. 44

circles on the left side appear farther apart than the unaffected circles on the right side. Consequently in the observation with Figure 45 *both* pairs of circles actually recede from the affected area. *Both* distances are "too large." But the

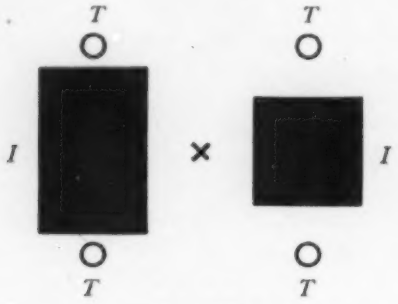


FIG. 45

displacement is greater on the right side where the *T*-objects are objectively farther distant from the *I*-object.

As a further instance we may mention Figure 46 which is in every respect analogous to Figure 31 except that in the present figure the *T*-circles lie just as much outside the parallel *I*-lines as they lay inside in Figure 31 (p. 289). After the inspec-

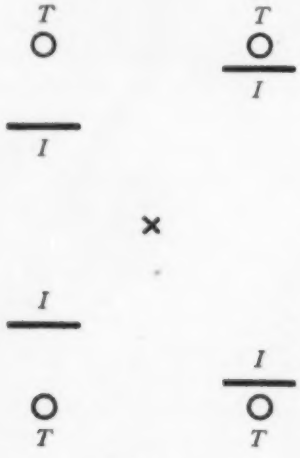


FIG. 46

tion period the distance on the left side appears greater.

Additional proof will be found in an observation with Figure 47. Here the *T*-objects are outline squares, actually shown on a white ground, which lie at equal height but at different distances from the contours of two *solid I*-oblongs. After the inspection period the *T*-objects appear at different heights. The square on the left side has a lower position than its partner. Both objects recede from the horizontal contours of the oblongs toward their less affected interior.

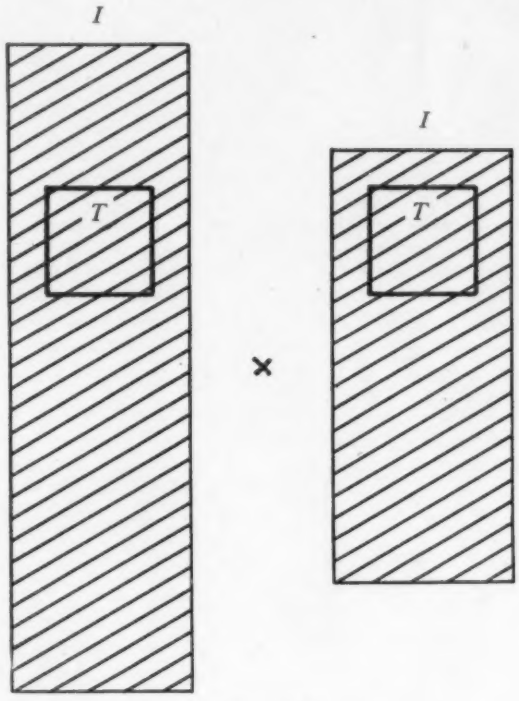


FIG. 47

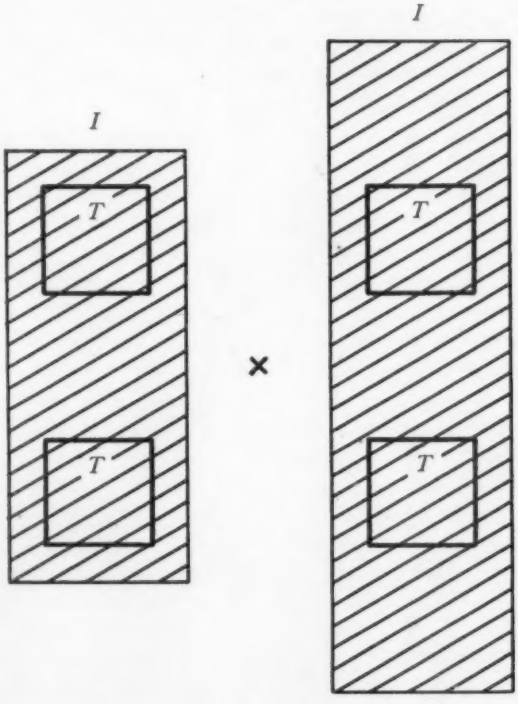


FIG. 48

But the square which is farther distant from the next contour does so more strongly. Once more the effect can of course be enhanced if four *T*-objects are used in a proper position. In Figure 48 the distance between the squares on the right side appears clearly shorter.

We have thus shown that the distance paradox is a general characteristic of displacements. It does not depend upon the presence of any particular size effects which may result when *T*-objects recede from satiated regions.

The principal phenomena which are described in this report can easily be verified by anyone who will spend a short time preparing the required drawings and a still shorter time repeating our observations. As a rule these observations need no corroboration by quantitative experiments. After all, when we demonstrate positive or negative after-images, color contrast, and the like, we also convince people without any measurements.

However, when the displacement of objects from satiated regions is said to be the most important fact in our field, a quantitative evaluation of this fact seems indicated. More particularly, it will be felt that the distance paradox needs stronger confirmation than is given by the semi-quantitative technique of our differential experiments. Obviously, it will be our task to measure the actual amount of displacements which occur at various distances from a satiated region, and thus to establish the paradox in strictly quantitative terms.

Some measurements of figural after-effects have been reported in Chapter Two. It will be remembered that in those experiments the method of constant stimuli was used. Single comparisons, each preceded by its inspection period, were given widely spaced in time because it was assumed that otherwise the after-effect might grow by accumulation of satiation.

This tedious procedure was not repeated in the measurements which we are now going to describe. Gibson has pointed out—and we agree with him—that with increasing inspection periods figural after-effects first grow rapidly and then more slowly until eventually a maximum is reached beyond which no further prolongation of the inspection time will lead to a further increase (*cf.* 9, p. 461). It seems that at this stage continued inspection makes for more persistent rather than for more intense after-effects. No particular inspection time can be given in which an after-effect becomes maximal. This time will depend upon the conditions under which the effect is established in each case. But if in a given instance the time of maximal effect is actually reached frequent repetitions of this inspection period will clearly no longer influence the size of the effect. Such repetitions will merely serve to keep that size constant. Therefore, in the following experiments the method of constant stimuli was used with this deviation from our former procedure: when after a given inspection period the position of the affected object had been compared with that of a neutral object, the next inspection period followed after about one minute, and so forth. We found that with this procedure results were only occasionally disturbed by accumulation when the inspection period was forty seconds. As soon as a trend did seem to develop, comparisons were discontinued until a sufficient rest period had elapsed. On a given day the displacement only in one particular situation was measured. Even under these conditions the gathering of judgments in sufficient numbers remained so cumbersome that with one subject the method of constant stimuli was replaced by the method of production. This variation of the procedure served also as a control.

Our *T*-objects were two outline squares which lay at equal distances on the left and on the right side of a fixation mark. On the right side—and in one experiment on both sides—an *I*-figure was first shown which affected the relative height of the two squares. The position of one *T*-square was varied in a random order, and the subjects reported in each case which square, if any, appeared higher. From a table of such judgments the most probable equivalence point was inferred.

Our apparatus consisted of two glass plates which were pressed against each other in a wooden frame. To the plate which faced the subject a white square (one inch square) with a black outline along its edges was attached. This plate was stationary. The plate behind it bore a second, and equal, square some horizontal distance apart from the first square. By means of a simple device this second plate could be moved up and down so that its square assumed any desired position with regard to the square on the stationary plate. Positions of the movable plate with its object could be read on a scale. Halfway between the *T*-squares a black dot, the fixation mark, was placed on the stationary plate. Immediately behind the glass plates a white screen was put. In our (constant) illumination the interior of the white *T*-squares appeared grayish against the white background, the square behind the front glass a

trifle more so than the other. Subjects were seated at a distance of about three yeards from the apparatus.

The accuracy of our measurements proved to be much greater than had been expected. While, as usual, our patterns were drawn in terms of inches, available scales in fractions of an inch were found to be too crude for our purpose. Thus, while the conditions of our experiments will be described in inches, the results will be given in millimeters and fractions of millimeters. Actually it seems improbable that the errors of our measurements ever exceeded one-third of a millimeter. The difference in height of the squares varied by steps of only one-half of a millimeter.

In a first experiment our intention was merely to gain an impression of the extent to which *T*-objects are displaced in the neighborhood of *I*-figures. We chose the situation in which one *T*-object recedes upward from its *I*-figure while a second *T*-object recedes downward from another *I*-figure (*cf.* Fig. 36, p. 292). The *I*-figures were solid black oblongs on a white cardboard. The *T*-squares were shown on the apparatus which has just been described. Their distance from the *I*-oblongs was one-quarter of an inch. From one comparison to another the left square was given varying heights. This means that on the right side the spatial relation between the *T*-object and the affected area remained constant whereas on the left side the *T*-square approached the affected area when it was lowered so as to compensate for the after-effects on the left and the right side. The two authors served as subjects. It was found that the squares appeared to have the same height when the right square actually lay higher

by 4.5 mm. with subject K., and
by 4.0 mm. with subject W.

When, with subject K., the method of production was used, seven determinations of the equivalence point gave as an average 4.5 mm., which is identical with the above value for the same subject.

These figures do not represent accurate measurements of the effect which results from the displacement upward on the left side and the displacement downward on the right side. For, when the left square is objectively shifted it moves not in a homogeneous medium but in a medium which is heterogeneously affected at various heights. None the less it remains true that in the two affected areas the two squares seemed to have the same height when one lay that much lower than the other. Figure 49 shows how two such squares look in normal vision when their objective heights differ as much as they did at the equivalence point of subject W.[12] They ought to be inspected from

FIG. 49

a distance of about three yards. Obviously, a person to whom these two squares would seem to have the same height *without* previous satiation would have to be considered as a patient. It was therefore not necessary to determine expressly at what relative position the two squares had the same phenomenal height in the absence of satiation. We know from other observations that under these circumstances neither of us would make an error of more than a small fraction of a millimeter. On the other hand, precisely during these measurements with their ever repeated inspection periods satiation sometimes became so persistent that after twenty-four hours the squares still appeared to be in a clearly oblique position when they had the same objective height.

If the left *I*-oblong is omitted only the right *T*-object is affected. It recedes downward from the remaining oblong. In this case any objective shift of the left *T*-square changes its position in a medium which we may regard as entirely unaffected. The position of the right square with regard to the affected area remains strictly unaltered. This means that in the present situation our procedure represents a measurement in the strict sense. With the authors as subjects the squares were now found to have the same phenomenal height when actually the right square lay

3.0 mm. higher with subject K., and
2.3 mm. higher with subject W.[13]

[12] Reduction as in the other Figures.

[13] With subject K. the displacement was also measured when in the same configuration the solid oblong was replaced by an outline oblong of equal size. Under these conditions the displacement was reduced from 3.0 to 2.3 mm.

In a further measurement the area of the solid oblong was reduced by four-fifths. Only a black stripe, corresponding to the lowest parts of the former *I*-object remained. As a consequence the displacement became

2.5 mm. for subject K., and
2.0 mm. for subject W.

If the differences are reliable we must assume that the displacement of *T*-objects depends not only upon their distance from the edge of the *I*-figure but also upon the area of this figure.

Up to this point the vertical distance between the *I*-object and the *T*-object was kept constant (one-quarter of an inch). Our measurements assume of course much more significance when they are not restricted to a constant distance. If we can rely on our differential experiments the distance paradox must appear as soon as the affected *T*-square is shown in varying positions relative to the *I*-oblong. In the following measurements nothing but this one factor was varied. The *I*-figure on the right side was the solid oblong in its full size. Four subjects took part in these measurements. Five distances were examined, but not all these with all four subjects. Results are given in table 1. For

TABLE 1

	Distances					
Subjects	2	3.5	5	6.5	8	inches/20
K.	2.2	2.0	3.0	2.0	1.8	mm.
B.[14]		1.6	2.8		2.1	mm.
W.	1.8	2.1	2.3	1.7		mm.
M.	.8	1.8	1.8	.5		mm.

reasons which have been mentioned the displacements are given in millimeters, the objective distances between the *I*-figure and the *T*-object are in inches. Translation of data from one unit to the other would have led to awkward numbers. Actually in the first line of table 1 one-twentieth of an inch is the unit of measurement. It will be seen that individual differences are as great as they are always found to be when figural after-effects are measured. But the distance paradox shows in the results of all four subjects. Somewhere in the neighborhood of one-quarter of an inch the displacements are greater than they are at shorter distances. The present data also answer another question: Our previous experiments left no doubt as to the fact that within certain limits displacements grow when *T*-objects lie at a greater distance from the *I*-object. Even a paradoxical fact, however, must remain within reasonable bounds. It was therefore to be expected that beyond a certain distance displacements would no longer grow. In fact, since at *very* great distances from an *I*-figure satiation cannot have any measurable effect it seemed a safe prediction that with increasing distance displacements would first grow and then again decrease. Our data show that this is actually the case. When the distance is greater than about one-quarter of an inch displacements are smaller than they are in this region, which thus appears as a zone of *maximal* effect.

The figures of our table are consistent with one exception. In the case of subject K. the displacement is a trifle too high for the shortest distance. Since with a given subject measurements at varying distances were sometimes separated by several weeks it would not have been surprising if more such instances had occurred. Elementary processes in the frog's nerve are known to depend upon seasonal factors. Thus, the brain of a human subject cannot be expected to remain exactly the same system irrespective of fatigue and variations as to food, weather, and the like.

Although these measurements referred to a special situation and particularly to satiation by a special *I*-object, we were at first inclined to draw from our results certain conclusions which eventually proved to be erroneous. When comparing outline figures with solid figures of the same size the former appeared to us as the "weaker" *I*-objects. Moreover, we assumed that because of this "weakness" outline figures satiate a smaller area than do solid figures. On these premises it seemed possible that the distance at which *T*-objects are maximally displaced is *shorter* when the *I*-object is an outline figure. We were willing to admit that the crucial distance might be the *same* with both kinds of *I*-objects.

Several experiments which were planned with this presupposition in mind failed completely.

[14] With this subject the method of production was used. It worked in his case as it has worked in many others; but it must be used with caution. Bales and Follansbee (1) report that some of their subjects saw no "curved line" effect. In actual fact the attempt to *measure* the effect gave the value zero. This does not prove the absence of the phenomenon since the authors worked with the method of production. We once used this technique with a subject who shows fairly conspicuous figural after-effects when other procedures are chosen. But with the method of production he was never satisfied by his achievements and corrected them so long that the final result was zero. Too much time elapsed, the after-effect as such decreased, and it was also obliterated by the prolonged satiating effects of the *T*-object in its various positions.

Their result was invariably the opposite of what we had expected. At last we decided to examine our premise that in the neighborhood of outline figures the distance of maximal displacement is equal or shorter than it is with solid *I*-figures. We found at once that with outline objects as *I*-figures the displacements of *T*-objects are maximal at a much *greater* distance than they are in the neighborhood of solid *I*-figures.

An estimate of the crucial distance was obtained in the following fashion. Two strong horizontal lines, each two inches long, were the *I*-figures. They were shown on both sides of the fixation mark which lay lower than the lines. The height of each line could be varied independently. Two outline squares served as *T*-objects (*cf.* Fig. 50). When the lines were

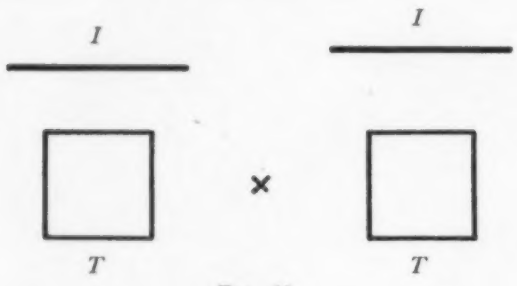


FIG. 50

put at varying, and different, heights above the *T*-squares either one square or the other was at a distance at which its displacement was greater. To be sure, we had to expect that in some cases the positions of the squares would not noticeably differ. But whenever one distance gave a stronger effect than the other then this distance lay nearer the distance of maximal effect.

Take the following list of judgments. Distances from the *I*-lines are given in terms of inches. Each judgment indicates at which of two distances the displacement downward was greater. It will be seen that the distance at

$$.6 > .4$$
$$.75 > .5$$
$$.7 > .55$$
$$.7 > .6$$
$$.75 > .6$$
$$.85 < .6$$

which the displacement is maximal must lie between .6 and .85 inches. In the same way the authors and a third observer judged differential effects for varying pairs of distances until it was entirely clear that with these three subjects the critical distance lay in the neighborhood of three-quarters of an inch. This is about three times the distance at which the displacement was found to be maximal with a solid *I*-figure. Of course, our result does not mean that at this greater distance the *amount* of the displacement is the same as it is at the much shorter optimal distance when a solid *I*-figure is used. In the latter case that amount may be greater. But upon this question our experiments have no bearing. They decide solely at what distance from an *I*-line a *T*-object is more strongly displaced than it is at other distances from the same line. The fact, however, that this critical distance is much greater with a mere line than with a solid *I*-figure will appear as puzzling. Actually it seems just as paradoxical as the distance paradox *per se*.

The critical distance depends not only upon the *I*-object which satiates the medium; it varies also with the *place* at which the displacement occurs: in the periphery of the visual field that distance is greater than it is near the fovea. In Figure 51 the horizontal distance between the *T*-object and the fixation mark is seven and a half inches. The *I*-objects are *solid* oblongs. In such places as have been used in previous experiments the critical distance would therefore be about one-quarter of an inch. On the right

FIG. 51

side of Figure 51 this is actually the distance between the *I*-figure and the *T*-squares while on the left side the corresponding distance is a full inch. Nevertheless after the inspection period the four squares appear clearly as a trapezoid having its longer side on the *left*. It follows that so far in the periphery the critical distance is greater than one-quarter of an inch even if the *I*-object is a solid figure.

When with the same *I*-figures the distances between these figures and the *T*-squares were as large as one inch and one and three-quarters inches one subject saw no clear distortion of the *T*-pattern; two others found again that the *T*-squares which lay farther apart from their *I*-figure were also farther apart from each other. We are inclined to conclude from these observations that, measured in retinal coordinates, satiation spreads wider in the periphery of the visual field than it does in parafoveal regions.

Quite apart from their intrinsic interest, such observations show that the distance paradox cannot be explained by eye movements which may occur during the inspection period. One might be tempted to say that such eye movements enlarge the area which is actually satiated, and that the distance paradox is somehow caused by this fact. We doubt whether this can be called an explanation of the paradox. At any rate, both with the arrangement of our quantitative experiments and with the present patterns the eye movements in question would have to be in the vertical dimension. There is no reason why in this dimension their excursion should be greater when the *I*-objects lie farther to the right and to the left. It is also quite unlikely that eye movements which occur during the inspection period will ever have the amplitude which that explanation would have to assume in the present instance.

III

The implications of given facts will sometimes not be clearly realized until these facts are stripped of any special and unnecessary ingredients. This applies to several observations which have been described in section II of Chapter Two. It was then found that within an angle which served as an *I*-object *T*-squares converged toward the open side of that angle. At the time this fact was interpreted in terms of a size effect which was, paradoxically, stronger at a greater distance from the sides of the *I*-angle. In the meantime it has been recognized that in the last analysis such size effects invariably involve displacements: outlines and outside parts of *T*-objects recede from affected areas. Consequently, if size effects show a paradoxical behavior they do so because this behavior is a general characteristic of such displacements.

This principle applies to the distortion of a square within an angle. Near the apex of an angle which has a horizontal position the horizontal outlines of a square approach the sides of the angle so much that only a small displacement results. Toward the opening of the angle the distance between its sides and the outlines of the square grows. Consequently the displacement of the horizontal lines away from the angle and toward each other must also grow, and the total effect will be that the square turns into a trapezoid which converges toward the open side of the *I*-figure. In section I of this chapter the same interpretation has been applied to the fact that within an angle the sizes of two separate objects or two vertical distances show the distance paradox. At present we merely extend this interpretation to instances in which a *continuous* variation of the distance between *I*-lines and *T*-lines causes a *continuous* variation of displacements.

We have previously made it clear that our interpretation ought not to be understood in an atomistic sense. We now repeat that at a given point the displacement does not depend merely on the distance which separates the *I*- and the *T*-line just at that point. This is impossible simply because the degree of satiation opposite a certain point of the *T*-outline is not determined by a single point of the *I*-line, which may, geometrically, be regarded as opposite. Rather, satiation in a given region is a function of the *I*-configuration which extends in the neighborhood of this region (*cf.* p. 274). Moreover, *T*-figures are much too coherent for any independent displacements of their "points" to be seriously considered. On the other hand, it remains true that in a molar sense conditions which make for displacements of *T*-outlines are least favorable near the apex and most favorable near the open side of the angle, while intermediate conditions in continuous gradation prevail between the two extremes.

In the preceding section it has been shown that the distance at which the displacement reaches its maximum is about three-quarters of an inch when the *I*-figure is a line. In the experiments now under discussion the distances between the sides of the angle and the outlines of the *T*-square did not exceed that critical amount. Correspondingly we never found that the displacement of the *T*-outlines near the open side of the angle was smaller than it was at places near the apex.

It is a matter of some interest that in receding from the sides of an angle the horizontal outlines of a *T*-square remained *approximately* straight lines. In our former description of these experiments some exceptions have been mentioned (p. 287). In fact, we observe quite frequently that under such conditions straight lines do not remain perfectly straight. It is nevertheless true that generally their curving is not at all strong. We are not convinced that this fact is due only to a tendency of straight lines to remain straight when any outside factors tend to distort their simple form. This tendency as such is a demonstrable fact. But in our connection it does not suffice as an explanation because under certain conditions straight lines are known to yield to the distorting influence of satiation. Rather it seems that in the neighborhood of a straight *I*-line the displacement of a *T*-line grows more or less as a linear function of the distance between the two.

As in a previous discussion, the present interpretation of after-effects within an angle may not seem to be more than a repetition in other words of what has been said in Chapter Two. Actually, however, our present formulation in terms of displacement calls attention to new experimental possibilities. For instance, *all* straight *T*-lines must suffer the same displacement if their position relative to the sides of an angle is properly chosen. Moreover, the interior of an *I*-angle no longer appears as the only place in which lines will turn into an oblique position.

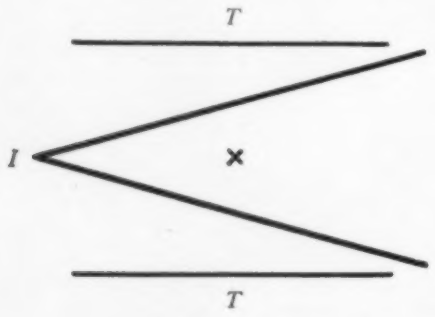


FIG. 52

With the arrangement of Figure 52 both inferences can be tested at the same time, because the *T*-pattern here consists of two separate parallels which both lie *outside* the angle. If our reasoning is correct these lines must recede outward from the sides of the angle, and they must do so more strongly where they are farther apart from the angle, i. e., on the side of the apex. After the inspection period the lines did diverge toward this side. Incidentally, in this particular instance the *T*-lines did not appear as altogether straight. Rather on the right side they formed extremely flat curves with concave sides facing the angle. It may also be mentioned that the lines were not symmetrically affected: the displacement of the lower line was clearly stronger (*cf.* p. 281).

At this stage no more than two steps remained to be taken if our tests were to assume the simplest possible form. In the first place, a *single* *T*-line was placed in a position in which it could be expected to turn in the same fashion, while another line, parallel to the first, lay farther off in the same direction as a neutral comparison object. Secondly, however, we had to assume that the former line would also turn into an oblique position if the more distant side of the angle was omitted. Our expectation was verified with the pattern shown in Figure 53.

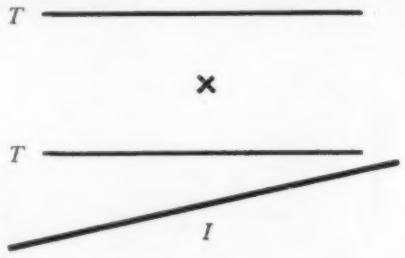


FIG. 53

After the inspection period the lower of the horizontal lines appeared turned upward on the left side. As a deviation from the horizontal position this could have been observed even if the more distant horizontal line had not been present. But this comparison object, which remained perfectly horizontal, served to make the obliqueness of the lower line more conspicuous.

Now, only the fact that this experiment was derived from observations with an *I*-angle caused us to regard a single *I*-line as one side of an angle. Apart from this accidental connection our *I*-object is simply a straight line, and the affected *T*-object is a second straight line which forms with the first a small acute angle. Thus the outcome of the present experiment establishes an important rule: ***After inspection of a straight line, other straight lines in its neighborhood, which form a small angle with the I-line, turn in space so that the angle grows.***

This rule covers not only the present instance to which it applies directly but also the distortions observed within an *I*-angle, which now appear as mere complications of the same simple fact. It will be realized, however, that the rule cannot be regarded as an independent and irreducible principle of figural after-effects. For it is a more basic principle that *T*-objects recede from satiated areas, and that distance affects such displacements in the paradoxical fashion. The principle of displacement can be established without any reference to lines which turn in the neighborhood of satiated regions. On the other hand, once that principle is established the turning of lines in certain situations can be predicted. Conversely, if it were known only that in such situations *T*-lines turn away from *I*-lines, the general principle of displacement could not safely be deduced from such knowledge, because the oblique position of the *T*-lines with regard to the *I*-lines might be an integral condition of the fact that the *T*-lines turn. Therefore, though our rule about the turning of lines in the neighborhood of *I*-lines is correct and extremely useful, the principle of displacement with its distance paradox is the more general and fundamental principle. To be sure, it remains possible that the *amount* by which a *T*-line turns in the neighborhood of an *I*-line is influenced by secondary factors peculiar to this situation.

Gibson, it is true, would insist that, just as in our experiments within an *I*-angle, it is not *any* line but a line inclined with regard to the *horizontal* which here makes another line turn. However, in the principle of displacement which suggested our present experiment there is no reference to any particular direction in space. We have no reason to assume that *T*-lines will turn only if the *I*-lines are given in special positions. In fact, evidence which will soon be presented makes such an assumption untenable.

With regard to our own rule we must, however, once more emphasize that its application to more complicated configurations requires some caution. It will be remembered that the turning of a *T*-line is not to be interpreted as the result of many independent displacements in which the various "points" of that *T*-line recede from opposite "points" of the *I*-line according to their local distance. Similarly, distortions in more complicated patterns are not simply the sum of displacements which the various lines of a *T*-pattern suffer in the neighborhood of individual *I*-lines. "Complicated" situations are not mere aggregates of simpler situations such as the one mentioned in our rule about the turning of *T*-lines. For instance, the rule applies to *T*-lines within an *I*-angle. But in applying it we must realize that satiation within an angle is not the juxtaposition of two satiation effects which the two sides of the angle cause in separate areas. There is no reason why satiation near the end of a simple straight *I*-line should be particularly intense. However, when two lines intersect, and thus form an angle, satiation must be much more intense near the apex of that angle than it is at points near its open side. To this extent, therefore, the turning of lines within an angle will occur under conditions which differ from those prevailing near a single *I*-line. And of this difference we must take account when we apply our rule to the interior of an angle.

Incidentally, this remark applies also to the *shrinking* of *T*-objects within an angle. The distribution of satiation within an angle depends upon the presence of its *two* sides. Under these circumstances it is clearly not quite correct to speak of size effects within an angle as though they consisted of two independent components: one displacement of one line away from the upper side of the angle and a second displacement of another line from its lower side. In our former analysis the actual situation has therefore undoubtedly been simplified. But no harm has been done. For the necessary correction can easily be made. Moreover, this correction does not lead to any essentially new facts. And, most important of all, the simple situation to which the rule refers is by no means a fiction. On the contrary, that situation has just been realized in our observation with Figure 53.

From our simple rule we can derive further facts. Some such facts will be discussed in the next chapter. At present one example will suffice.

From previous observations (*cf.* p. 286) we know that within a *closed* *I*-object, also, *T*-lines turn when they are shown in a proper position. But if an oblique *T*-line turns whether it is given inside or outside an *I*-figure, such a line must also turn when it lies half inside and half outside that figure. Moreover, on both sides of the boundary the direction of its turning will be the same; because the angle between the boundary and the *T*-line tends to grow both inside and outside.

This situation is realized in Figure 54. After inspection of the trapezoid it will be found that the affected *T*-line is actually turned, and that the direction of its displacement agrees with our prediction. Comparison with the second *T*-line

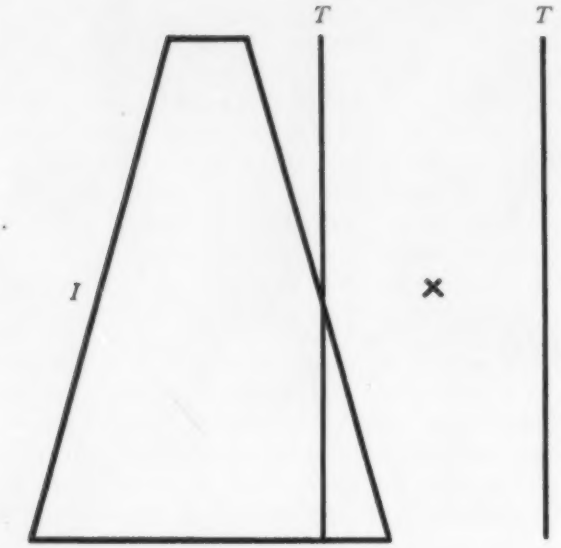


FIG. 54

serves to make the obliqueness of the affected line more conspicuous. Many observers report that within the area of the *I*-object the line deviates more strongly from the vertical than it does outside; in other words, that the line does not remain quite straight. This is not surprising. Quite apart from the anisotropy of visual space to which we have repeatedly referred, satiation just inside the boundary is stronger than it is just outside (*cf.* p. 273).

A more impressive effect will be obtained if two objectively parallel lines intersect the two oblique sides of the *I*-object (Fig. 55). The two lines turn in opposite directions so that their convergence downward is fairly striking. This observation may also be repeated with a solid *I*-object.

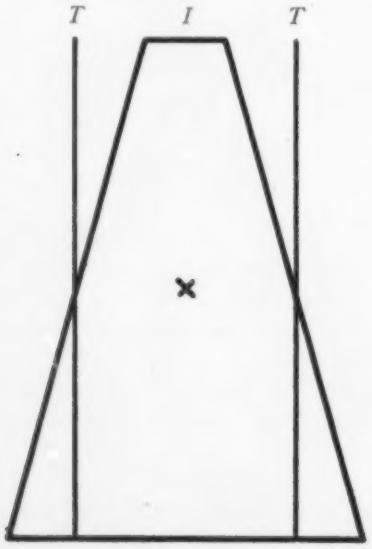


FIG. 55

We have not yet found a subject who fails to see the convergence of the *T*-lines in this pattern. Moreover, for the most part very short inspection periods seem to suffice. To the authors the effect is always noticeable after an inspection time of no more than five seconds; although with such short inspection periods often it is only the interior half of the lines which will appear displaced.

CHAPTER FOUR

GIBSON'S DISCOVERIES

We are now prepared to consider Gibson's discoveries within the general descriptive scheme which has developed from our own observations. This task offers no particular difficulties because the after-effects which have been described by Gibson derive from our rule about the turning of lines in the neighborhood of satiated regions. There is only one instance to which this statement does not apply.

I

1. The rule about the growth of small angles between an *I*-line and a *T*-line applies directly to Gibson's "bent line" effect. This term, it will be remembered, refers to the fact that previous inspection of a bent line transforms a straight *T*-line into a line bent in the opposite direction—provided, of course, that both lines have approximately the same location and orientation.

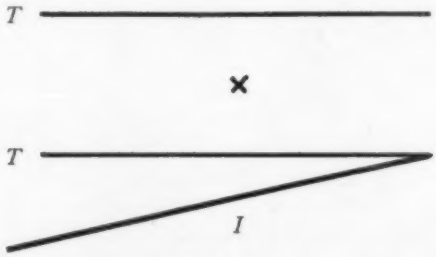


FIG. 56

Now, if in an arrangement like that of Figure 53 the oblique *I*-line and the *T*-lines are continued until the *I*-line meets the lower *T*-line (Fig. 56), the resulting configuration is objec-

tively just one-half of the arrangement with which Gibson demonstrates the "bent line" effect (Fig. 57). It does not matter that the higher parallel of Figure 56 is not contained in Gibson's arrangement. We have used this line merely as a standard of reference. It has surely no essential influence upon the turning of the affected line, and it can be omitted if this dis-

FIG. 57

placement is discernible without its presence. As one must expect, observation with the pattern of Figure 56 gives once more the result with which we are familiar: the lower parallel is clearly turned upward on the left side.

From this it follows that in the pattern of Figure 57 the horizontal *T*-line cannot remain a straight line. Rather its two halves will both recede from the corresponding halves of the bent *I*-line, and as a result the *T*-line will be seen as bent in the opposite direction. But, the present pattern is a complete replica of Gibson's "bent line" arrangement. We have therefore shown that the "bent line" effect is deducible from our scheme of concepts: the rule about the turning of lines and, in the last analysis, the principle of displacement with its distance paradox. The fact that in Gibson's pattern the halves of a straight line turn in opposite directions makes his experiment a particularly striking demonstration of that principle once the principle is known. Human vision is highly sensitive to deviations of a line from straightness.

It is not a serious objection to this interpretation of the "bent line" effect that Gibson's arrangement represents a "complicated" situation, and that such a situation is not entirely reducible to simpler patterns (*cf.* p. 305). To be sure, inspection of a bent line does not cause quite the same satiation as would be caused by two straight lines, the halves of the bent line, taken separately. The bent line represents an angle and, although this angle is very wide, satiation near its apex must be a trifle more intense than it is elsewhere along the bent line. If we had to take account of details we should therefore have to make a certain correction. But this correction would only mean that there is probably a small quantitative difference between the whole "bent line" effect and the combination of two simple "turning" effects.

2. We shall now turn to the discovery which first called Gibson's attention to the existence of certain figural after-effects, the "curved line" effect. After inspection of an arc with a great radius the objectively straight tangent to the middle of the arc appears curved in the opposite direction. Figure 58, a modified form of Figure 56, will show at once that this effect, too, can be deduced from the rule which has been formulated in section III of the third chapter. The *I*-line is now one-half of Gibson's curve. There is no reason why the lower *T*-line should not recede from this curve if it recedes from the straight line in Figure 56. As a matter of fact, it will be found that in this pattern the lower parallel does curve upward on the left side. (The second

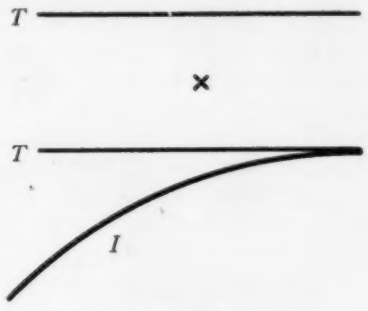


FIG. 58

horizontal line is of course no longer needed. The curving of the affected line is obvious without any comparison object.) Suppose now that Figure 58 be completed toward the right so that we obtain Gibson's arrangement. From the present observation we can predict that in this larger pattern the horizontal *T*-line will curve upward at both ends. An adequate interpretation of Gibson's discovery will therefore be obtained if we can show why in one-half of the present pattern the straight *T*-line assumes a *curved* shape when it recedes from the *I*-curve. Actually this is a direct consequence of the fact that with a *straight* *I*-line the *T*-line remains approximately straight as it turns away from the

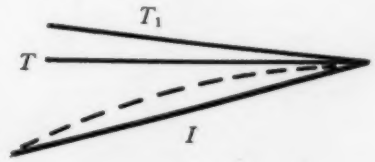


FIG. 59

I-line. For the sake of simplicity we shall assume that in the latter case the *T*-line shows *no* curvature at all. In our next figure (Fig. 59)

T_1 represents the T-line in its displaced position. If in this position (T_1) the T-line is still a straight line its displacement at any place is obviously proportional to the objective distance between the I-line and the T-line at this place. We now replace the straight I-line by the corresponding arc of a circle—the broken curve of Figure 59—while the T-line remains objectively unchanged. Let us assume that in receding from this arc the T-line follows about the same rule as has just been found to hold for the case of the straight I-line, namely, that along the T-line displacements are about proportional to the objective distance between the I-curve and the T-line. It will be seen that, if we begin on the right side this distance is at first smaller than it was with the straight I-line, but that its increase toward the left side is *accelerated*, until eventually the distance of the T-line from the I-curve equals its distance from the straight I-line. Under these conditions the T-line must show very little displacement on the right side. As we gradually proceed to the left, however, the displacement must be accelerated; it must grow by ever increasing amounts from the right to the left. Now we know that without such an acceleration of displacements the T-line remains approximately straight. Consequently the acceleration of displacements must make the T-line curve upward. The same will happen at both ends of the T-line if this line and the I-curve are completed toward the right so that Gibson's arrangement results. Thus Gibson's "curved line" effect follows from our rule and from the fact that with a straight I-line the T-line remains more or less straight when it turns.

We have not tried to show that, in the "curved line" effect, the T-line must turn into part of a circle when the I-line is such an arc. Any such attempt would have been unwarranted because for the most part the affected T-line appears by no means as part of a circle. If, for instance, the I-curve is shown in a vertical orientation the lower half of the T-line will not only tend to curve earlier but also more strongly than the upper half (*cf.* p. 281). Moreover, depending upon the particular curvature of the I-line and on its length, the T-line will sometimes be curved mainly at its ends, sometimes more strongly at its middle. In our observations such deviations of the T-line from an exact circular arc have been the rule rather than an exception. We must therefore limit our discussion to the fact that in the neighborhood of an I-arc a straight T-line appears as *some* curve which is convex toward the I-arc.

3. Strictly speaking, our interpretation of the "bent line" effect and of the "curved line" effect cannot be regarded as final until we know that Gibson's third discovery, the "tilted line" effect, also fits into our descriptive scheme. In this respect our situation is as follows:

There can be no disagreement concerning the *fact* which Gibson has called the "tilted line" effect. If a straight I-line is given at a small angle relative to the vertical (or horizontal), an objectively vertical (or horizontal) T-line will assume an oblique position. The direction of its obliqueness will be opposite to that of the I-line. Gibson's interpretation of this fact is implicit both in the formulation which has just been given and in the term "tilted line" effect. In other words, reference is made to the vertical and the horizontal not merely because under the conditions of Gibson's experiments the I-line happened to form a small angle with one main axis of space or the other. Rather it is assumed that small deviations from the vertical or horizontal actually *cause* the effect. On this ground an interpretation of the effect in which the main axes of space are never mentioned will obviously seem to miss the crucial point. Also, on the same ground, the effect must disappear when its cause is absent, i. e., when the I-line is given in orientations in which it does not form a small angle with the vertical or the horizontal. Gibson would therefore say that it is impossible to derive the "tilted line" effect from any general rule about the turning of lines near satiated regions, irrespective of the orientation of these lines. Besides, he would add, the existence of such a rule cannot be inferred from our experiments; because in all observations which have been adduced for its support, moderately oblique I-lines have been combined with actually vertical or horizontal T-lines. Hence these observations must be regarded as mere verifications of the "tilted line" effect.

Nor would our critic stop at this point. He would extend his objections to our interpretation of the "bent line" effect and the "curved line" effect. These after-effects, he would say, can be demonstrated in any orientation of the bent or curved I-lines and their respective T-lines. Thus, since we have not actually shown that the growth of small angles between I- and T-lines is independent of the orientation of these lines, no valid interpretation of the

"bent line" and the "curved line" effects has yet been given.

From an experimental point of view it comes then to this. Does satiation *always* increase small angles between an *I*-line and a *T*-line? Or is this phenomenon restricted to instances in which such an increase constitutes a "tilted line" effect in Gibson's sense? In the next section we shall describe observations which will give a clear answer. It must, however, be repeated that such experiments will decide not merely about the existence of a special "tilted line" effect. Rather, in the same observations our interpretation of the "bent line" effect and the "curved line" effect will be put to a final test, because this interpretation is valid only if the "tilted line" effect is absorbed within a general rule about the increase of small angles by satiation.

II

A preliminary answer to our question can be given by simple qualitative experiments. For instance, an *I*-angle in which a square is placed as a *T*-object may be turned until the pattern no longer represents a "tilted line" situation. Similarly, a trapezoid the oblique lines of which are intersected by two parallel *T*-lines (*cf.* Fig. 55) may be given in orientations which are quite arbitrarily chosen. It will then be found that the square and the pair of parallel lines still show the distortion which has been described in previous sections.

However, the best qualitative test is as follows: In our former Figure 56 (p. 306) a straight *I*-line which deviated by 12.5° from the horizontal was combined with two parallel *T*-lines which were actually horizontal. The lower horizontal had one end in common with the *I*-line; the other horizontal which lay at a greater distance served as a neutral standard of comparison. In this situation the fact that the former line turned upward on its "free" side could be interpreted as a "tilted line" effect. But this interpretation does not apply if the whole pattern is turned by 12.5° so that the *I*-line itself assumes the horizontal position, and the two *T*-lines are now tilted by 12.5° (*cf.* Fig. 60). From Gibson's point of view no "tilted line" effect can occur under these conditions because the *I*-line does not deviate from the horizontal "norm." Actually the outcome of the experiment is not perceptibly affected. The lower horizontal turns away from the *I*-line just as it does with the arrangement of Figure 56.

But a crucial issue deserves a more accurate treatment than is provided by merely qualitative observations. Hence we decided to measure the growth of small angles between an *I*-line and a *T*-line when these lines were given in various orientations.

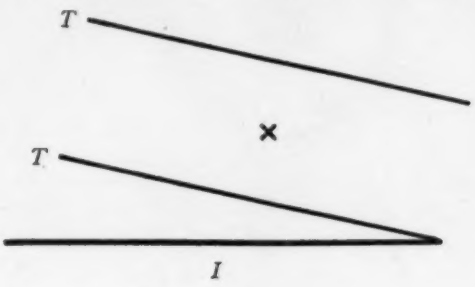


FIG. 60

The principle of our measurements was quite simple. An *I*-line was inspected in a given orientation, and then a *T*-line was shown in another orientation so that a small angle lay between the two lines. Parallel to this *T*-line there were on both sides and at some distance two other lines which served as comparison objects. The position of the *T*-line in the middle was not actually constant. As one inspection period followed another this line was shown in slightly varying positions, and in each case the subject compared its direction with that of the fixed lines. The method of constant stimuli was used throughout.

More concretely, our arrangement was like this. In a large white cardboard screen of circular shape a rectangular opening was cut with edges 4 and $3\frac{13}{16}$ inches long. Against this screen a second white cardboard circle was pressed so that the center of this circle coincided with the center of the opening in the screen. Therefore a black *T*-line (3 inches long) which was drawn through the center of the second circle passed also through the center of the opening in the first circle. By turning the cardboard circle with the *T*-line the orientation of this line relative to the edges of the stationary opening could be varied. Subjects were able to recognize minute deviations from the position in which the line in the middle and two edges of the opening were parallel. Their judgment was further facilitated by the fact that on the stationary screen black lines similar to the *T*-line were drawn along those edges. (*Cf.* Fig. 61). Objectively the orientation of the *T*-line with regard to the edges was measured in terms of angles by which the circle with the *T*-line was turned against the screen with the opening.

On a third white cardboard circle and through its center the *I*-line was drawn $3\frac{1}{2}$ inches long. Observers fixated first the middle of this line for forty seconds, and then the center of the *T*-line.

After one inspection period the *T*-line was shown in only one particular orientation relative to the opening. When the subject had judged, the next inspection period followed after a rest period of two minutes, and so forth. In this fashion judgments were collected for a number of positions until the point of apparent parallelism could be safely inferred from the data. In this respect the procedure was the same as the one previously used. There was, however, one difference. With our present technique the *T*-line itself was shown in varying positions while the neutral comparison lines remained stationary. Strictly speaking,

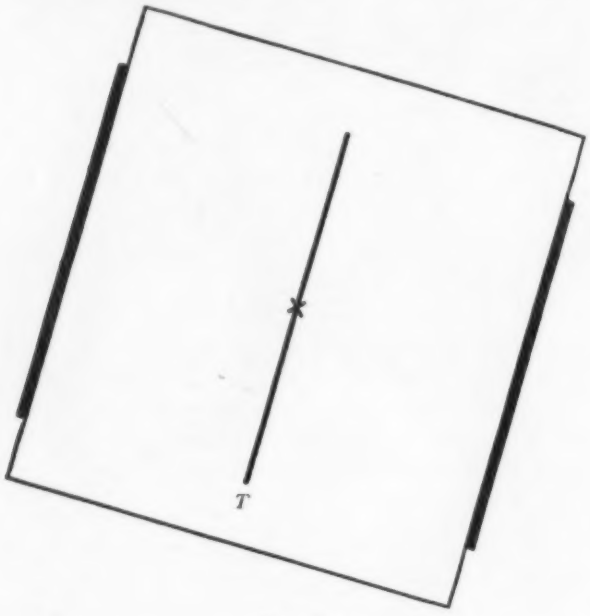


FIG. 61

in a measurement of figural after-effects the neutral comparison objects ought to be varied because only in this case will variations and measurements occur in a homogeneous medium. When varying the position of the affected object one moves it in a medium which previous fixation of the *I*-object has made inhomogeneous. Under these circumstances results are perhaps not entirely exact measurements of after-effects. It would have been possible to turn the comparison lines and to leave the *T*-line stationary. But we decided to use the technically simpler procedure which more closely resembles Gibson's measurements. His subjects compensated for the "tilted line" effect by turning the *T*-line.

In these experiments the rest periods, after each inspection period and the subsequent judgment, proved to be indispensable. When they were omitted a disturbing trend in the distribution of judgments soon developed. Obviously with mere lines as *I*-objects an inspection period of forty seconds does not yet cause a maximal effect, and if such periods follow one another too rapidly the after-effect begins to grow by summation. In fact, even with interpolated rest periods it is hardly safe to have a subject judge more than ten times in succession.

In table 2 the orientation both of the *I*-line and the *T*-line is indicated in the following

TABLE 2

I = orientation of the *I*-line; *T* = orientation in which the *T*-line is objectively parallel to the comparison lines

		1	2	3	4	5	6
	I	10°	0°	35°	100°	85°	90°
	T	0°	10°	45°	90°	95°	100°
Subjects	W.	1°30′	1°15′	2°56′	1°10′	1°20′	1°26′
	K.	1°15′	1°15′	2°52′	1°26′	1°15′	1°37′
	N.	54′	59′	2°20′	1°26′	1°47′	1°21′

fashion. To a horizontal line the angle 0° is ascribed. The vertical position is indicated as 90°; and if the upper end of the line lies on the left side of the vertical the size of the angle is given as greater than 90°. In each experiment the positions of the *T*-line were varied in steps of 21′. Although these steps were given in a random sequence judgments proved to be highly consistent. Subjects observed from a distance of two yards.

Table 2 contains the results for three subjects and for six positions of the *I*-line and the *T*-line. In each case the orientation of the neutral comparison lines differed from that of the *I*-line by 10°. And the results indicate by what angle the *T*-line had to deviate from the direction of these standards if the *T*-line was to appear parallel to them, i. e., if compensation for the after-effect was to be obtained. Invariably the *T*-line had to be turned toward the direction of the *I*-line; in other words, satiation always *increased* the angle between the *I*-line and the *T*-line.

The measurements in columns 1 and 4 are analogous to measurements reported by Gibson and Radner (9). In one case the *I*-line deviates 10° from the horizontal position, in the other case by the same angle from the vertical position. The results indicate by what angle the *T*-line had to be turned toward the *I*-line if it was to appear horizontal (1) or vertical (4). The effects have about the size which could be expected from the work of Gibson and Radner. These authors used *I*-lines which formed an angle of 5° either with the horizontal or the vertical directions.

From columns 2, 3, 5, and 6, however, it seems perfectly clear that such after-effects have no special relation to the horizontal and vertical positions. In fact, this table could not possibly suggest such a relation. In the first place, the after-effect is not smaller but, on the contrary, much greater when both the *I*-line and the *T*-line deviate by large angles from the horizontal and the vertical (3). In the second place, with the *I*-line in either the horizontal or the vertical

position (2 and 6), the effects are on the average just as great as they are when that line forms a small angle with the main axes of space. According to Gibson's view there should be no after-effects under such circumstances because no tilted lines are given during inspection. It might be suggested that Gibson's term "tilted" refers to deviations from the phenomenal rather than the objective horizontal and vertical. But when for this reason we planned to repeat the experiments 2 and 6 with the *I*-line in positions which were phenomenally vertical or horizontal for individual subjects, we discovered that for one of the subjects of table 2 the subjective vertical showed no reliable difference from the objective vertical; and that for another subject the same held with regard to the horizontal. This made any further experimentation unnecessary.[15]

To the measurements with an angle of 10° between the *I*-line and the standard for the *T*-line we added a few experiments in which that angle was 20°. Apart from this difference the procedure was in every respect the same. Subjects K. and W., the authors, were observers in these measurements, which gave the results contained in table 3.

TABLE 3

		1	2	3
	I	−10°[16]	35°	80°
	T	10°	55°	100°
Subjects	W.	48′	1°26′	47′
	K.	1°48′	2°31′	1°40′

The essential fact in this table is that once more a stronger after-effect is obtained in the neighborhood of 45° than near the horizontal or the vertical. In this sense the correlation of the effect with the principal axes of space is negative. If to this finding we add the fact that a horizontal and a vertical *I*-line cause after-effects just as do *I*-lines which are tilted we come to the conclusion that *a special "tilted line" effect does not exist.* A *T*-line which forms a small angle with an *I*-line turns away from the latter in whatever orientation the pair of lines may be shown. We cannot yet say for what reason this effect is greater in strongly oblique than in slightly tilted positions.

When Gibson measured what he called the "tilted line" effect he found it particularly *small* (or absent) when the *I*-line had the orientation 45°. There is no contradiction between this observation and the fact that in our experiments the after-effects are particularly strong in the neighborhood of 45°. Gibson varied the size of the angle between the *I*-line and the *T*-line while the *T*-line remained either horizontal or vertical. Under these circumstances the after-effect tended to disappear when the angle *between the lines* approached 45°. From this nothing follows concerning the size of after-effects when both lines together are shown in various orientations while the angle between them remains small and constant.

During our quantitative experiments we had ample opportunity to repeat an observation which has been mentioned before. Although it is quite true that the judgments of the subjects were highly consistent, in the beginning each subject had to overcome a certain difficulty. Without previous satiation it is perfectly easy to compare the direction of the *T*-line with that of the standards. In this a slight change occurs when the *I*-line has been inspected, inasmuch as now the affected line is often not strictly straight. Apart from other minute distortions which are sometimes observed the *T*-line appears almost regularly as a line which is just a trifle bent. In other words, its two halves are not turned by exactly the same amount. But subjects tend to become reconciled with this fact and to pass judgment on what may be called the general direction of the line. At any rate, measurements are not seriously disturbed by this minor irregularity.

It will be admitted that in view of this evidence the existence of a "tilted line" effect can no longer be maintained. The experiments which once seemed to establish such an effect do not prove more than that a tilted *I*-line causes the same effect as is caused by straight *I*-lines in any other orientations. At the same time, of course, our data show that our interpretation of the "bent line"—and the "curved line"—effects is fully justified. If small angles between an *I*-pattern and a *T*-line grow in all orientations of these lines, such effects *must* exist.

Gibson's main discoveries have just been shown to fit into our scheme of concepts. Only one fact is left to which these concepts do not apply. This fact is Gibson's "indirect" effect. He found that if an *I*-line formed a small angle with the vertical (horizontal), it turned not only a vertical (horizontal) *T*-line but also a horizontal (vertical) *T*-line. With the *I*-line at 80°, for instance, a horizontal *T*-line turned in such a way as to *decrease* the large acute angle between the two.

When we tried to repeat this observation under the conditions which we had used in measuring the growth of small angles we never obtained a

[15] The fact that in table 2 certain results occur repeatedly has no particular significance. The repetition of those results is brought about by our procedure which does not permit more than crude interpolations in reference to the steps actually presented to the subjects.

[16] In the orientation −10° the right end of the *I*-line lies below the horizontal.

measurable effect. But, to judge from his report, Gibson's measurements of the "indirect" effect were not made under such conditions.

Although he was aware of the fact that the "direct" effect persisted when at some distance the objective vertical and horizontal were indicated by neutral lines and edges, he had at the time decided to exclude such landmarks. His subjects observed through a cone so that the straight experimental lines appeared alone in its circular opening.

When we used a similar procedure our subjects did observe the "indirect" effect. But it again disappeared when objective verticals and horizontals were shown somewhere in the environment. In view of this difference between the "direct" and the "indirect" effects we hesitate to believe that the latter phenomenon is a figural after-effect in the sense in which we are using this term.

There are, however, other observations to which the "indirect" effect may be related. In 1912 Wertheimer (29, pp. 97 ff.) found that if by means of a mirror the content of a whole room was given an oblique orientation it tended after a while to appear in a normal or comparatively normal orientation. This fact, which we will call "normalization," has been confirmed by G. G. Brown (2) who turned the environment by means of prisms, and more recently by Kleint (13). It is true, Gibson and Mowrer (11) report that they were not successful when trying to repeat such experiments. But if we are to judge from their description of the facts the basis for this negative statement seems to be merely that visual space did not assume an *entirely* normal orientation. Surely, this is a bit beside the point.[17] About the fact of *approximated* normalization there seems to be little doubt.

In two respects normalization appears akin to figural after-effects. With many subjects it tends to develop *gradually* like these effects, particularly if the visual content of space is turned by large angles. Moreover, normalization can be said to lie in the same direction as the turning of lines in figural after-effects. But such effects never turn a line by angles of a size comparable to the frequently huge effects of normalization. On the other hand, there is some evidence that normalization is not so stable as figural after-effects in our present meaning of the term.

We are not yet in a position to indicate what relation actually obtains between the two phenomena. It would be important to know to what extent normalization can be demonstrated under conditions in which a figural after-effect of the same direction is strictly excluded, for instance, in places which are so far distant from any tilted contours that figural after-effects are no longer possible. Naturally, in the future it will be more important to discover some connection between the two phenomena than to stress their difference.

Did normalization ever interfere with our own observations? We do not think so, because in our experiments the main axes of space always appeared stable in their normal orientation. But it seems quite possible that Gibson's "indirect" effect belongs in this category.

[17] The authors also quote Brown's paper as confirming their negative results. Actually Brown's report (*cf.* pp. 134 f.) emphasizes the convincing character of normalization.

III

Gibson sees his discoveries within a particular frame of reference. This scheme we must now briefly examine because our own observations have a direct bearing upon its main concepts, that of visual norms and that of figural adaptation. A straight line is a norm in comparison with curves and bent lines. Again, among the straight lines there are norms of a higher rank, the lines which lie in the vertical or the horizontal directions. According to Gibson figural after-effects result from the fact that a figure which represents a deviation from some norm undergoes a process of adaptation when inspection is prolonged. Adaptation makes the figure more similar to the norm than it objectively is. But the norm itself is also affected by this process. If after adaptation the norm is actually shown as a *T*-object it will appear altered in the sense in which the inspected object was changed by adaptation, and this distortion is the figural after-effect. Gibson concludes that the amount of the after-effect is equal to the amount of adaptation which has occurred in the *I*-object.

Against this explanatory scheme two objections must be raised. In the first place, we cannot admit that visual norms in Gibson's sense are of primary relevance to the occurrence of figural after-effects. Secondly, changes which an *I*-object suffers during prolonged inspection are as a rule not quantitatively equal to the after-effects which *T*-objects show afterwards.

1. No difference of opinion exists concerning the concept of visual norms as such. To be sure, straight lines and the main directions of space are not the only visual norms. We do not know whether Gibson assumes that there are no other such norms. Furthermore, Gibson's statements about norms occasionally show a slight empiristic trend of which we cannot approve.

Our present criticism refers to the statement that a deviation from such a norm is the essential cause of a figural after-effect. This statement is incompatible with the fact that inspection of a norm causes after-effects just as does inspection of any other figures. In section II it has been shown that this holds not only for

straight lines the direction of which has been arbitrarily chosen but also for vertical and horizontal lines (*cf.* also p. 277).

It should not be inferred that for this reason we deny any influence of visual norms upon figural after-effects. Such a view would be inconsistent with our conception of visual norms. If these norms are outstanding cases in visual experience, the figure processes which correspond to them must have particular characteristics in a physiological or physical sense. Therefore satiation by a "norm process" is also likely to differ in some respect from satiation by other figure processes. And by the same token it will make a difference whether or not the *T*-figure is a norm.

Perhaps the most essential difference between Gibson's theory and our own views is this. Gibson holds that in a latent form the characteristics of a figural after-effect are established when the *I*-figure has been inspected; in other words, that they are established independently of the *T*-object. Our observations prove that an after-effect depends just as much upon the *T*-figure which is then shown as upon a given pattern of satiation. It is the spatial *relation* of a *T*-figure to that pattern of satiation which determines the more important after-effects. More particularly, spatial *differences* between an *I*-object and a *T*-object are decisive. Roughly speaking, such differences can be translated into distances by which parts of the *T*-object are separated from parts of the *I*-object. And upon these distances depends the way in which the various parts of the *T*-object will be displaced.

On this ground, it will be seen, we may take any figure as our *T*-object and derive from it an *I*-object the inspection of which must cause a figural after-effect in that *T*-object. The *T*-object or the *I*-object or both may be norms. Again, neither may be a norm. With a proper choice of spatial relations between the two an after-effect will always result.

Suppose the *T*-object is actually a norm. We can then construct an *I*-object by changing the spatial characteristics of that *T*-object more or less *ad libitum*, but within certain quantitative limits. Under these conditions phenomena analogous to Gibson's discoveries follow with almost geometrical necessity. Take our observation with a trapezoid the oblique edges of which are intersected by two parallel *T*-lines (*cf.* p. 306). The divergent edges of the *I*-trapezoid may be regarded as deviating from a norm, namely, the parallel orientation of the edges. When this norm is actually given as a *T*-pattern we find it distorted. And we cannot be astonished by the fact that this distortion is just the opposite of the deviation from the norm in the *I*-figure. For, in choosing as our *I*-pattern a modification of the "parallel norm" we have introduced just those relations of position and of distance which make such a "mirror effect" necessary. In places in which the *I*-object is "too narrow" it lies by construction inside the *T*-object, which must therefore grow "too wide" as its lines recede from those of the *I*-object. Conversely, where the *I*-object is "too wide," the *T*-object lies inside so that when its lines recede from the *I*-object the *T*-object must become "too narrow."

It will suffice if we add merely two further examples. In Figure 62 the inner square of the

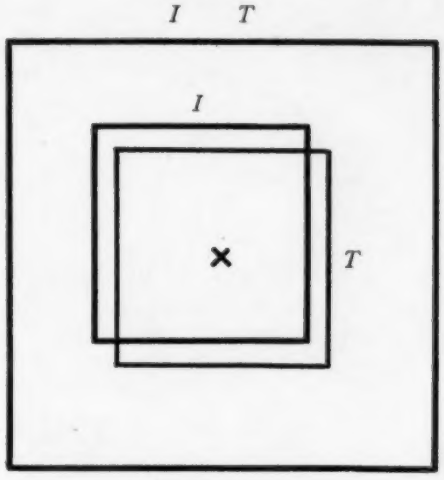


FIG. 62

I-pattern deviates from a concentric position which would give the pattern the character of a norm. That square lies too far to the left and too high. If the concentric pattern is given as a *T*-object the position of the inner square differs from its position in the *I*-pattern by certain amounts and in certain directions. The principle of displacement states that (within certain limits) these differences between the *I*- and the *T*-pattern will be enhanced when the *I*-pattern has been inspected. Hence the inner square in the *T*-pattern must be displaced from its concentric position, and the sense or direction of this eccentricity must be the opposite of the one which the *I*-pattern exhibits. This is exactly what happens when the experiment is actually

performed: the inner *T*-square lies too far to the right and too low.

Again, in Figure 63, where a circle is the norm and a certain ellipse is used as an *I*-figure, the circle must turn into a kind of ellipse the longer axis of which lies at right angles to that of the *I*-ellipse. In places in which the *I*-ellipse is "too long" the circle is smaller, lies inside, and must shrink when it recedes from the ellipse. Conversely, in places in which the *I*-ellipse is "too narrow," the circle lies outside, and must expand when it recedes from the ellipse. In Figure 63 the resultant displacements are indicated by arrows.

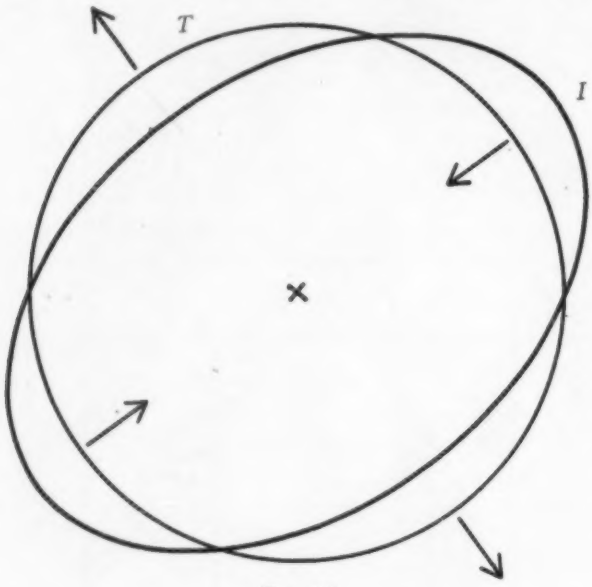


FIG. 63

In this fashion, with a norm as *T*-object and a modification of that norm as *I*-object, the norm will always show distortions which are complementary to the deviations of the *I*-object from the norm. This is simply a consequence of the way in which such an *I*-object is constructed from a given *T*-norm.

In view of Gibson's theory it must, however, be emphasized that a given figure may either be regarded as an *I*-object or as a *T*-object. Therefore, if a pair of figures is given, their roles will be interchangeable in this respect. Since distances between the two figures remain the same whether one or the other is chosen as the *I*-object we must expect that, in general, analogous displacements will occur in both cases, although their direction in the one case will be the opposite of that in the other case. Thus it can once more be shown that after inspection of a norm, too, other figures will be distorted—provided their form and location are properly chosen. For obvious reasons observation of the distortion may be more difficult under these circumstances than it is when the norm serves as the *T*-object and the other figure as the *I*-object. But the fact that such experiments *can* be performed must be regarded as crucial evidence against Gibson's theory.

2. We may now turn to Gibson's statement that during the inspection period an *I*-object changes as much by adaptation as the *T*-object is changed in the after-effect, and that both changes lie in the same direction. We have no serious objection to the term adaptation so long as adaptation is merely a name for such figural after-effects as occur when an *I*-object serves as its own *T*-object. Certain observations which have been mentioned in the beginning of our report (*cf.* p. 271) belong in this class. Moreover, so long as a norm serves as a *T*-object and one particular modification of the same norm as the *I*-object, Gibson's statement may appear fairly plausible. But after inspection of a given figure a great many different patterns may be used as *T*-objects. According to the nature of these *T*-objects and to their location relative to the *I*-pattern, various figural after-effects will be observed. Quite apart from differences as to shape, such distortions will also be *quantitatively* different. Therefore they cannot all be equal to the change which adaptation has caused in the given *I*-object.

That Gibson's theorem is not generally applicable can also be shown in the following way. According to his thesis not only the amount but also the direction of the after-effect is the same as that of the preceding adaptation. An *I*-curve, for instance, becomes less curved during the inspection period. A straight line which is then shown appears convex in the direction in which the *I*-curve became less curved. Take, however, a straight vertical line as an *I*-object. As a *T*-object we may choose a straight line which intersects the *I*-line at an angle of 10°. But actually *two* such lines exist: one with upper end lying on the left side of the vertical *I*-line and another with upper end lying on the right side. If the former is taken as the *T*-object it will turn farther to the left; but when the latter is chosen it will turn farther to the right. Thus these after-effects occur in opposite directions. Now, it would be difficult to imagine any dis-

tortion which could be caused by adaptation if the *I*-object is a straight vertical line. But apart from this difficulty adaptation of a given figure can surely not occur in opposite directions at the same time.

We have just said that it is difficult to imagine what distortion by adaptation can occur in a straight vertical line. With certain other *I*-objects we find ourselves in the same predicament. In Figure 64, for instance, eight angles are combined so as to form a star. If after the inspection period this figure is removed, and if now a *T*-square is put into the former place of an angle this square will be distorted in the familiar fashion. But how can that symmetrical *I*-figure be distorted by adaptation? A change in the size of a particular angle would mean that its neighbors must change in the opposite sense. Quite apart from the symmetry of the figure which argues against such an assumption, the square may be put into any angle, and it will always be distorted in the same way. Consequently, if there is any distortion by adaptation it must be the same in all angles. Since this is impossible in our figure, Gibson's thesis can evidently not hold in the present instance.

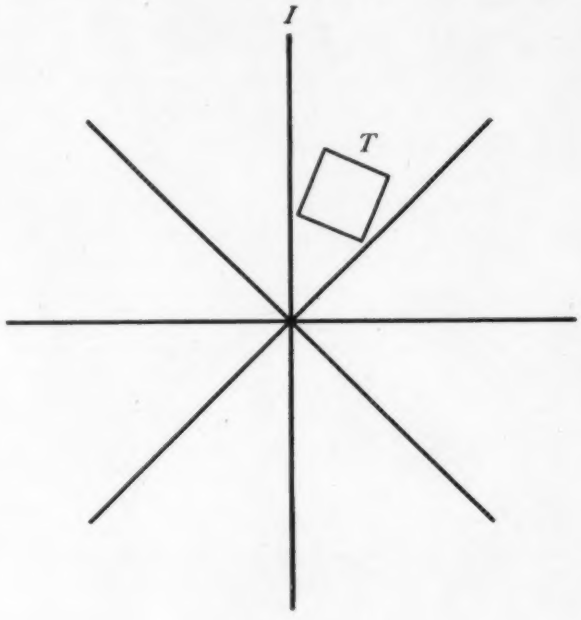


FIG. 64

In those figures in which adaptation does occur, we often find it less conspicuous than are the figural after-effects which properly chosen *T*-objects show after inspection of those figures. In fact, it was this quantitative difference which in Chapter Two called our attention to the distance paradox. We do not wish to deny, however, that with certain patterns adaptation causes fairly striking effects. Some such instances will be discussed in another chapter.

Gibson reports that in certain instances he measured the amounts both of adaptation and of after-effect, and that they were equal. It seems to us that his experiments do not prove his point. For, if we read correctly, the amount of adaptation was not directly measured. Rather it was inferred from certain observations which are open to another interpretation. There seems to be only one procedure by which the amount of adaptation can be actually measured. This procedure is the direct comparison of the adapted *I*-object with varying objects of the same class which are shown in an unaffected region. It seems that Gibson did not perform such measurements.

To summarize: It has been shown that in an interpretation of figural after-effects the concepts of visual norms and of distortion by adaptation cannot play a decisive part. On the other hand, Gibson's main discoveries fit into a descriptive scheme which has developed in the present investigation.

CHAPTER FIVE

DIRECT CURRENTS AND THEIR AFTER-EFFECTS

In Chapter One it has been stated that inspection of any figure, of any specific entity in a visual field, gives rise to figural after-effects if test objects are properly chosen. Chapters Two to Five contain additional support for this general statement. In fact, if it were doubted that the statement holds for a particular object we could at once indicate tests in which figural after-effects *must* appear after that object has been inspected. For, all specific things in visual space share some basic characteristics with the *I*-figures which have been used in our experiments. And the rules which describe the dependence of the after-effects upon given *I*-objects refer to such basic characteristics rather than to any particularities which merely individual figures possess.

On the other hand, *only* things, figures, lines, patches, dots, i.e., particular visual entities, cause such after-effects. In every statement which formulates our observations the presence of such specific entities is presupposed. Moreover, the neutral areas in which we show unaffected *T*-objects are as a rule white surfaces in which the intensity of sensory processes in

the usual sense must be great. None the less, in the absence of things, figures and the like, the presence of those intense sensory processes during the inspection period establishes no appreciable satiation and therefore no figural after-effects. Thus, from a functional point of view specific visual entities must be characterized by processes of their own which cause satiation and corresponding after-effects.

From this we can draw a conclusion the relevance of which transcends the special field of the present investigation. It seems no longer possible to interpret the existence of visual things, figures, or other particular entities in visual fields as a result of learning processes. According to such interpretations those specific entities have no truly visual existence. It is only past experiences which give certain fractions of a visual field the character of specific things, figures, lines, and so forth. But it will not be denied that in our experiments prolonged inspection of figures affects the medium in which they are psychophysically represented; in other words, that satiation spreads within and around those particular parts of the visual brain in which the figures happen to be physiologically located. From those empiristic theories it would therefore follow that learning processes of the past are capable of causing satiation and figural after-effects in the visual brain. It will take an intrepid theorist to defend such an extraordinary view.

Our argument is not weakened by experiments in which familiarity with certain figures is shown to have an influence upon the perception of other patterns. The very technique of such experiments presupposes the existence of those figures. They are either known beforehand or made especially familiar as given figures, and as such figures they influence the perception of patterns which are subsequently shown. Consequently such experiments do not refer to the present issue. This issue concerns the alleged formation of figures from a material which is supposedly devoid of anything like figures (*cf.* 12, p. 21).

I

Elsewhere it has been suggested that any segregated entity in the visual field is associated with an electric current in the visual sector of the nervous system, that this "figure current" polarizes the tissue through which it flows, and that figural after-effects are observed when later the current of a *T*-figure spreads in the affected medium (16, pp. 213 ff.; 17, pp. 73 ff.). The present study of figural after-effects has greatly strengthened our confidence in this assumption. In the following chapter we shall see that the assumption explains not only the existence of figural after-effects as such, but also their seemingly paradoxical behavior. We therefore propose now to discuss the functional characteristics of electric currents to the extent to which these characteristics have a bearing upon our subject.[18] Not all currents can interest us. We are not concerned with beams of electrons or of other charged particles, which move like bullets through a rarefied atmosphere. The currents which are relevant to our purpose flow under tremendous friction within solid or fluid conductors. More particularly, we shall consider *direct* currents, i. e., currents the flow of which constitutes a steady state.

1. In a steady state a current is a functional unit in this sense: the flow at a given place is being maintained by the fact that correlated amounts of current flow elsewhere in the circuit, and vice versa. More specifically, the current constitutes a steady state when the *same* amount of flow passes through all cross-sections of the circuit. Now, the term "same" indicates a *relation* among the amounts of flow at various cross-sections. We can therefore say that a steady current is *relationally maintained.* For this reason particular attention was given to electric currents when it was first recognized that steady physical states exhibit the principal characteristics of *Gestalten* (15, pp. 133 ff.).

The fact that a steady current is relationally maintained has immediate consequences as to the way in which the distribution of the current depends upon the properties of its conducting medium. Take Ohm's law, and apply it once to a particular conductor within a closed circuit and then to the circuit as a whole. In a conducting stretch n, which contains no electromotive force and has the resistance r_n, the current is maintained by a difference of potentials ΔP_n between the ends of the conductor.

[18] We do not assume that all characteristics of electric flow are equally relevant to psychophysical theory. Their magnetic fields, for instance, are not likely to play an important part in the functions of the nervous system. This is the meaning of a previous statement by the senior author according to which electromagnetism and electrodynamics do not belong to the fields about which the psychophysicist must have expert knowledge (17, p. 115). The terms electromagnetism and electrodynamics refer only to such effects of electric currents as are mediated by their magnetic fields. Apparently this has not been recognized by W. R. Sickles and G. W. Hartmann who in a recent paper (24, p. 405) interpret K.'s statement as though it rejected K.'s own theory of figure currents.

We propose to compute ΔP_n in terms of given properties of the conducting circuit. Whatever the amount of ΔP_n may be, Ohm's law states that in the conductor n

$$\frac{\Delta P_n}{r_n} = i,$$

where i is the intensity of the current. But the same law applies to the circuit as a whole if ΔP_n is replaced by the total electromotive force E which maintains the current in the circuit, and the resistance r_n by the sum of all resistances (Σr) which lie in series within the circuit. Thus we have

$$\frac{E}{\Sigma r} = i.$$

Since in both formulas i means the same amount of flow we obtain

$$\Delta P_n = \frac{E \cdot r_n}{\Sigma r}.$$

This expression shows that in a steady current differences of potential are "justly rationed." The sum of all differences of potential along the circuit necessarily equals E. Any particular conductor n receives a share ΔP_n corresponding to the *ratio* $(r_n/\Sigma r)$ with which the resistance in this conductor contributes to the total resistance of the circuit. In other words, the distribution of potentials along a given circuit *depends relationally* upon the various resistances of that circuit. As a first consequence, the difference of potential between the ends of a given conductor is altered if resistances in *any* parts of the circuit are changed. As a second consequence, however, no changes of resistance in a given conductor affect the distribution of potentials if resistances in all other conductors are varied in the same proportion. The fraction $r_n/\Sigma r$ is invariant with regard to such changes.

This is only one among many examples of "transposing" in physics. In all such instances it is, of course, relational determination which makes "transposing" possible. In spite of this fact quite a few psychologists are still adverse to thinking in terms of relational determination. Instances of such determination are but grudgingly admitted as though it were an evil, and considerable energy is being spent on complicated theories which serve to keep the evil in narrow bounds. And yet relational determination is a common fact in elementary physics. Hence, there is no reason why its occurrence in psychology should disturb anybody.

If more than one path is open to a current its distribution among the parallel conductors again follows Ohm's law. It is well known that this distribution is also relationally determined. Suppose that at a certain point a the circuit divides into two branches which then again unite at a second point b. If i_1 and i_2 are the intensities of the current in the two branches, and if r_1 and r_2 are the corresponding resistances, then Ohm's law states that

$$\frac{i_1}{i_2} = \frac{r_2}{r_1}.$$

The two parallel branches taken together offer a "compound" resistance which is a simple function of their separate resistances. Obviously, if in one branch the resistance is increased the compound resistance also grows. But in a current, we remember, differences of potential between two points are proportional to the resistance which must be overcome between these points. Thus, if the resistance in one branch rises, and if, therefore, the compound resistance of both is increased, the difference of potential between the points a and b must grow. In one branch, however, the resistance has not been changed. Therefore, in this branch an increased difference of potential now drives a current through the same resistance as before. As a consequence the current in the unaffected branch must flow with increased intensity when the intensity in the other branch is lowered by increased resistance. This holds for any number of parallel branches among which a current may be distributed. From branches the resistance of which is increased the current will partly be withdrawn. At the same time the flow will become more intense in branches which have not been affected. Our next considerations are merely applications of this simple principle to more interesting circuits.

2. Resistances within a circuit may be altered either by influences from without or by effects which the current itself has upon its medium, the conductors. We are concerned merely with the latter alternative. A current may alter its own medium mainly in two different ways.[19] In the *first* place, heat production, which invariably accompanies the electric flow in a resisting me-

[19] There are more such ways. But those which we omit have no significance for our purposes.

dium, may raise the temperature of the conductors. The resistance of metallic conductors grows under these circumstances. *Secondly*—but this holds only for conduction in electrolytes—the current may polarize surfaces through which the charged particles within the flow cannot freely pass.

The first possibility is never realized in living tissue because there are no metallic conductors in organisms. We shall nevertheless consider a model of precisely this type in order to introduce a mode of thinking which will afterwards prove to be more generally applicable. We shall do so because most people are more familiar with the behavior of currents in wires than with other forms of conduction.

How does a current spread in a space-like and quasi-homogeneous medium? Let us imagine that a dense and regular network be constructed in which thin wires run parallel to the three axes of a co-ordinate system. Distances between parallel wires are supposed to be small and everywhere the same. The wires, which have throughout the same resistance per unit of length, connect at the corners of small cubes. We suppose that the network as a whole extends very far in all three dimensions.

Parts of this network we assume to be removable like units of a maze. We can therefore put a strong battery somewhere in the interior of the network and then re-establish the complete conducting system around the battery. With the interior of the battery we are at present not particularly concerned. As poles we give it two parallel metal plates of equal size and shape. Numerous wires of the network are in contact with these pole-plates. Figure 65 shows a cross-section of the system as it appears in the plane of one pole-plate. The oblong in the middle represents one pole-plate, and the points on it indicate the places at which wires join the plate at right angles to the plane of the drawing. The second plate with its corresponding points of contact is not visible because it lies just underneath the one which is shown in the drawing. The network, we repeat, is supposed to extend much farther than it does in Figure 65.

When the battery is inserted the self-distribution of electric flow will immediately be establish both in the battery and within the network. In the network the current will issue from one pole-plate; it will then bend in all directions around the battery and return to it by the other plate. In other words, the battery will be surrounded by a flow which passes through the wires of the network. Obviously, however, the intensity of the current will be greater along such conductors as offer a shorter connection of the pole-plates than along others which are parts of wider detours with their greater total resistance.

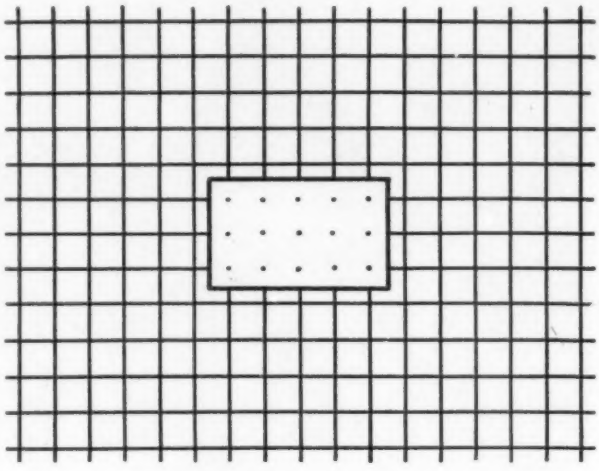


FIG. 65

Outside the battery the current will therefore be concentrated in the neighborhood of this source, and at a very great distance its intensity will be negligible. This follows directly from Ohm's law as applied to many conductors of varying total resistance which connect the poles of a battery.

As the current passes through the wires it heats these wires. If the battery is strong enough this effect will raise the temperature of the wires to such an extent as to increase their resistance. However, this change of the medium by the current itself will not affect all parts of the network to the same degree. The resistance will grow most in those wires in which the intensity of the flow and therefore the production of heat is particularly great, and the resistance will practically not be altered where the flow is fairly weak. Now, in the network of our model the intensity of the flow is greater in the neighborhood of the battery than it is at greater distances. It is therefore in the neighborhood of the battery that the resistance must grow most strongly. In more remote wires resistances will not be appreciably altered even though the current may spread to such distant wires.

Essentially, then, the situation is this. In the beginning the presence of the battery in the network is functionally represented by the elec-

tric flow which surrounds the battery. After a while the battery is surrounded not only by this flow but also by a zone of heightened resistance. If the current is now discontinued there will still be this affected zone, and it will remain until the temperature of the wires and with it their resistance approach their normal value.

3. Before the normal state of the network is re-established after-effects can be demonstrated within and around the affected zone. We may, for instance, insert a test battery in a place which lies between the most strongly affected region and a zone which has remained practically unaffected. Under these circumstances the distribution of the test current will follow the same rules as govern the distribution of all currents. Ohm's law remains the basic principle. But this law must now be applied to a network the various parts of which no longer have the same specific conductivity. Thus, since resistances are heightened in certain conductors in which a considerable part of the flow would normally spread, the current as a whole will be weakened. At the same time, however, the distribution of the test current will also be *distorted*. As has been explained on page 317 this current will partly be withdrawn from the region in which the network is strongly affected, and it will flow with increased intensity into regions in which the conductors have remained unaltered.

Clearly, this reasoning applies not only to a test current but also to the current by which the network is first affected. As resistances in the neighborhood of the battery grow this current is weakened by its own action. From this it follows that the modifying effect of the current upon its own substratum must tend to limit itself. The more the resistances grow the less current will flow through the affected conductors. Gradually, therefore, the current will become incapable of further raising the resistance in wires of which the resistance was first most rapidly altered. But just like the test current the first current is not only weakened. Its distribution is also altered. For, as the resistance in certain parts of the network grows, the flow must partly shift into other paths in which it was originally weaker.

4. In living tissue, which is an electrolytical conductor, the heat produced by a current has not the same effect upon conductivity as it has in metallic conductors. Moreover, under normal conditions currents which spread in the tissue of a human being do not alter the temperature of this medium to an extent which is worthy of note in the present connection. We must therefore modify our model and study the action of currents upon their conducting medium under circumstances which more directly resemble organic conditions.

Ohm's law applies no less to electrolytical conduction than it does to conduction in metals. The distribution of currents in electrolytes is therefore also relationally determined. In electrolytical conduction, however, a current may affect its own substratum in a way which does not occur in metallic conduction, namely, by polarization of the electrodes through which the flow enters and leaves the electrolyte. Since in electrolytes ions are the carriers of the current, certain ions must here accumulate while the concentration of others is lowered. These

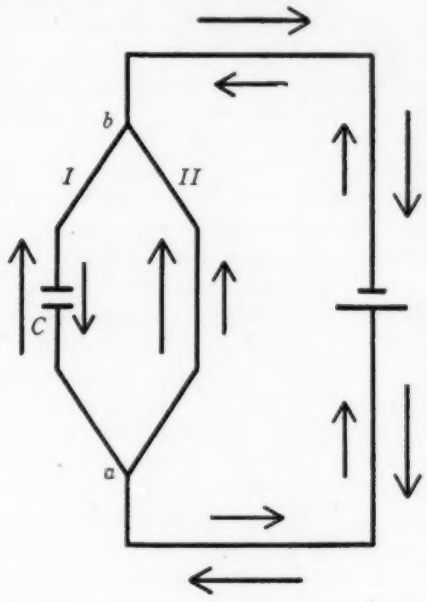


FIG. 66

changes at the surfaces of electrodes establish new electromotive forces rather than new resistances. But, since the direction of the new forces is opposed to that of the polarizing current, the result of their action upon this current is roughly the same as though the resistance at the electrode had been considerably raised. It will help us to understand certain phenomena in electrophysiology if we demonstrate this fact in a simple instance.

In Figure 66 *C* represents an electrolytical cell the electrodes of which are polarized by the cur-

rent of a battery. The cell is located in one of two branches *I* and *II* into which the circuit of the battery divides between the points *a* and *b*. After polarization the circuit contains two electromotive forces, that of the battery and that which has been established by polarization of the cell. The actual flow which under these circumstances passes through the given circuit is the sum of the currents which each electromotive force maintains independently. Obviously, in branch *I* the direction of the current caused by polarization of the cell is opposite to the direction of the current which issues from the battery. At point *a* the former current divides into two parts, one which flows through the battery to point *b* and thus back into the cell, and a second part which reaches *b* through branch *II*. It will be seen that not only in the cell but also in the battery the current caused by polarization flows against the current of the battery. On the other hand, in branch *II* both currents have the same direction, and therefore the flow in this branch is strengthened. (In Figure 66 the large arrows indicate the direction of the current which issues from the battery; the small arrows give the direction of the current caused by polarization of the cell). To this extent the effect of polarization is the same as though the resistance of the cell in branch *I* had been raised: from the affected branch *I* current is withdrawn, but in the unaffected branch *II* the current flows with increased intensity (*cf.* p. 317). On the other hand, there is, of course, an essential difference between polarization and an actual increase of resistance. If resistance is actually raised in a branch of the circuit, this change affects currents independently of their direction. The same does not hold for polarization, since polarization means that new electromotive forces are established which operate in a certain direction. If, therefore, after polarization the current of the battery is reversed, the effect of previous polarization upon this current is at first just the opposite of the effect indicated in Figure 66.

Apart from this particular fact, polarization in a ***network*** of conductors must obviously cause phenomena like those which follow a local heating of conductors by action of the current. Suppose that in all branches of the network electrolytical cells are inserted. The electrodes of these cells will be polarized by the current of the battery. As a result the battery will presently be surrounded by a zone in which the network opposes the passage of the polarizing current. When this current is discontinued, after-effects will appear as soon as a test battery is inserted. If the current of this battery has the same direction as the polarizing current, its distribution will be distorted in the same fashion as it would be by an increase of resistances in the same zone.

We can now approximate by one more step conditions which obtain in the organism. After all, it was merely for propaedeutic reasons that we chose to consider currents which spread in networks rather than in continuous volume conductors. In instances in which relational determination cannot be denied, the human mind still feels more secure if at least some remnant of separation is maintained. But, as a matter of fact, it makes little difference if we replace the network of our model by a conductor in which electric flow in any direction is possible at all points. Thus an electrolyte of practically unlimited extension may surround a battery. This battery will again be pervaded and surrounded by its own current, and on the outside the density of the current will be strongest in the immediate neighborhood of its source, because here the "lines of flow" are shorter, and resistances correspondingly lower. Now suppose that membranes of some kind are densely distributed throughout the conducting electrolyte. Such membranes are known to act more or less like metallic electrodes in an electrolytical cell. They prevent free passage of ions, and are therefore being polarized as soon as the current is established. In this fashion the battery surrounds itself with a zone in which new electromotive forces oppose the current of the source. Afterwards the currents of test batteries will be similarly affected, and therefore distorted, if their direction is the same as that of the polarizing current.

II

Nerve tissue is a volume conductor in which only electrolytical conduction is possible. We shall discuss, first, the alterations of this medium by currents from outside sources and, secondly, the currents which may originate in the tissue itself. Such currents, of course, cause the same changes in the tissue as do currents which issue from outside sources. (From our present discussion we shall exclude the fact that both at the beginning of their flow and upon their interruption electric currents tend to start nerve impulses.)

When two points of a nerve are suitably connected with the poles of a battery the phenomenon of *electrotonus* is observed. Electrotonus was discovered about one hundred years ago. The name derives from theoretical views which are now no longer held; but a considerable number of facts to which the name refers are firmly established. In the following pages we shall mention only those which seem relevant to our present problems.

Immediately after its onset the current begins to weaken. At the same time its distribution is changed inasmuch as a considerable flow appears outside the stretch which offers the most direct connection of the electrodes. The current seems to withdraw from those parts of the nerve's surface which lie directly under, or very near, the electrodes and to prefer a detour around these places. Among the observations which are covered by the term electrotonus this is the central fact. Incidentally, if a current passes through layers of many cells, for instance, through the skin, similar phenomena arise. Obviously the flow affects its living conductor so as to cause obstruction. This change is maximal in the places in which the current has the greatest density, i.e., under the electrodes, and therefore the lines of flow bend around these places.

Attempts have been made to interpret electrotonus in terms of polarization. Actually, as conductors, the surfaces of living cells in their environment of tissue fluid have certain properties in common with metallic electrodes through which a current passes into or from an electrolyte. It is also true that at those living surfaces polarization by direct currents occurs in a way which differs only in details from the polarization of metallic electrodes in ordinary solutions. Nevertheless, electrotonus in general cannot be identified with polarization because according to present experimental evidence electrotonus consists of two different effects of the current. A first and rapid effect *is* strictly comparable with electrolytical polarization. Where the current enters a nerve, under the anode, positive ions are accumulated on the outside and negative ions on the inside of the cell membrane. Just the opposite occurs where the current emerges from a nerve, i.e., under the cathode. To be sure, quite apart from electrotonus, positive charges are spread along the outside of the nerve's surface, and negative charges along its inside. But this "normal polarization" of the cell's surface is a condition along the surface to which electrotonic polarization seems to be simply added.

The change of charges both under the anode and under the cathode deflects the flow of the polarizing current from its most direct course. To the degree to which electrotonus is polarization this is to be expected. Several inanimate models have been constructed in which polarizing currents establish exactly the same phenomenon (*cf.* also pp. 319 f.).

Like electrolytical polarization in an ordinary circuit, the early stages of electrotonus may appear comparable to an increase in resistance. Actually, however, it must be remembered that new forces are being established which are opposed to those of the polarizing current, weaken this current, and strengthen the flow along detours. Any effects which at this stage electrotonus has on test currents will therefore vary with the direction of these currents.

The theory of electrotonus in this first sense represents one of the most satisfactory applications of physics to physiological facts. For our present purposes, however, other, more gradual, effects of electric currents on the tissue seem to be more important. In its early stages, we said, electrotonus is polarization. Here the term "early" is to be understood in an extreme sense, because it has been found that, just as physical theory demands, the polarization of cell surfaces is generally completed within a few milliseconds. If after this extraordinarily short period cell surfaces are still further altered by the current, such effects cannot simply be interpreted as additional polarization. Further alterations do, in fact, occur. They continue and grow not during tiny fractions of a second but for full seconds and minutes. Numerous investigations have dealt with these changes, but their physical nature has not yet been entirely clarified. Under these circumstances we propose to mention merely those facts which appear to be definitely established, and to restrict theoretical considerations to a few essential remarks.

During its early stages electrotonus develops symmetrically under and around the two electrodes. This symmetry is soon lost if the polarizing current continues to flow. The deviation of the current near the anode—the anelectrotonus—now becomes much more conspicuous, while the deflection near the cathode—the catelectrotonus—is now weakened. Moreover, as the anelectrotonus reaches a higher level it also spreads more widely in a spatial sense. With

this goes a further change. When the anelectrotonic effect has been firmly established in a given region not only those currents will be strongly deflected in this region which have the same direction as the polarizing current, but test currents which flow in the opposite direction will also show enhanced electrotonic deviation. Similarly, as the catelectrotonus gradually weakens, not only the polarizing current but also test currents which have the opposite direction are now less deflected from their most direct course. The explanation which is most commonly given ascribes these effects of continued electric flow to alterations of the very make-up of the cell membrane. In the anodic region this membrane is being "tightened," in the cathodic region it assumes a progressively "looser" consistency. The cell surface is never entirely impermeable. Consequently the degree to which a current accumulates ions on this surface will depend upon the consistency of the membrane. If the membrane is tightened it will be more strongly polarized, and if its consistency is weakened, the affected region will become less polarizable. For this reason the electrotonus which develops under the influence of continued currents must obviously be distinguished from the electrotonus, or polarization, which a current establishes immediately (*cf.* 4, p. 782). Polarization, it is true, is held to be essential in both cases. But after prolonged flow of a current the *primary* fact in electrotonus is not polarization as such but a change of the degree to which the cell surface obstructs the movements of ions. Because of such changes polarization can now go much farther in the anodic region and will become "subnormal" in the cathodic region. It is also these changes of polarizability which now affect test currents irrespective of their direction. Gradually, it seems, the anelectrotonic condition prevails while the catelectrotonus now plays a relatively minor role. When this situation has been established currents are apparently deflected in whichever direction they flow. Considerable electrotonic effects may therefore result even if *alternating* currents pass through the tissue. Naturally such effects are greatly reduced when the alternations have a high frequency.

This, we hope, is an adequate survey of the principal data concerning electrotonus, although some facts like the influence of electrotonus upon nerve impulses have been omitted. Inferences of our own we have tried to avoid entirely, and the explanatory principles which have been mentioned are those proposed by physiologists like Höber, Gildemeister, and Ebbecke (*cf.* 4).

Electrotonus is a condition which may persist for a long time. If a current of sufficient intensity has passed through the tissue for several minutes and is then discontinued, a high degree of electrotonus may still be demonstrable after fifteen minutes, and weaker effects still later. We are not acquainted with any special measurements in which the gradual disappearance of electrotonus has been followed to its end. There is no question, however, that the duration of electrotonic after-effects has about the same order of magnitude as the duration of visual satiation when inspection times have been sufficiently long.

III

We shall next discuss the nature of electromotive forces which arise in the activities of the nervous system itself. Our problem is whether there are any such forces besides those which cause the weak electric flow of nerve impulses.

Normal activity of the nervous system is still held to be a function mainly of individual neurons. According to many theorists activity of the nervous system means that neurons are thrown into action one after another in a longitudinal sequence. The fact that transmission at a synapse often seems to require the arrival of impulses in more than one fiber does not change this picture in any essential respect. Nor does it mean a fundamentally new principle if in several neurons which are arranged in a circular fashion impulses may repeatedly travel around the closed chain.

In his great report on central synaptic transmission Lorente de Nó (20, p. 439) defends precisely this interpretation of function in the nervous system:

> In view of the evidence now available there can be no doubt that in a first approximation it is permissible to explain the activity of the nervous system in terms of the activity of chains of neurons and the detonator actions of the nerve impulses, i.e., in terms of all the anatomical elements that the nervous system contains and of the nerve impulses that they produce or carry.

The author admits that the conception expressed in this statement may have to be corrected in the future. We are convinced that such a correction is unavoidable even now because many activities of the nervous system are relationally determined in a way which we can-

not understand in terms of separate actions within the anatomical elements.

Relational determination as such is not a concept to which physiologists will be inclined to object. For, the origin of any nerve impulse, i.e., of action in an anatomical element, is an instance of relational determination. No such impulse is possible until conditions in adjacent parts of a given fiber are made to *differ* by a local affection, the stimulation. Moreover, once the impulse has been established, its very nature as a current involves relational determination. Its spread in and around the fiber follows the principles explained on pp. 316 f.

However, a discussion will arise if it is assumed that, quite apart from impulses within individual neurons, specific processes spread across the tissue as a continuum when conditions in one part of the tissue become sufficiently different from those in an adjacent part. It is our contention that activities of the nervous system cannot be fully understood unless we adopt the concept of relational determination in this *molar* sense.

We need not go far if we wish to support this thesis by factual evidence. Such evidence is contained in our present investigation. Let us return to the beginning of this chapter where the same argument was used for a psychological purpose. A typical observation of figural after-effects has this form: in the left half of the visual field a homogeneous white surface is interrupted by a dark outline while in the right part the white ground is not so interrupted. Physiologically speaking, the retina is strongly stimulated in all places upon which parts of the white surface are projected. It is less strongly stimulated within the image of the dark outline. From those strongly stimulated parts frequent impulses in many fibers will issue; from the image of the outline less frequent impulses will travel in fewer fibers. Intense neural and synaptic activity will therefore be caused in those parts of the brain in which impulses arrive "from the white." Much less activity will be caused in parts which correspond to the retinal image of the dark outline. This is, in terms of neurons, impulses, and synaptic activities, the situation in the left half of the visual field, which contains the outline: function of anatomical elements on this side is weak at least in certain places. In the right half of the field where *all* impulses issue "from white" nervous activity is of course intense throughout. And yet, if after a while we introduce equal test objects in both halves of the field, the tissue on the left side proves to be profoundly affected while the right half appears to be in a comparatively normal condition. Clearly, so far as the activities of anatomical elements are concerned, it is the right half of the field which one would expect to be on the whole more strongly affected because in this half there have been more such activities than in the left half. None the less, just the opposite is true. Should we assume that given neurons and interneurons develop particularly hectic activities when action in their neighbors is at a low level? In the first place, no factual evidence seems to support such an assumption. In the second, if the assumption is made, activities of the nervous system are no longer explained in terms of chains of anatomical elements and their impulses. Rather, events in certain chains of neurons are now supposed to be relationally determined by the state of their neighbors. Moreover, since from a given outline satiation extends across considerable distances, impulses in parallel chains of neurons would have to be transversely interrelated across corresponding stretches in the tissue. We do not wish to imply that this is an impossible hypothesis; but we do maintain that the hypothesis, far from saving the customary interpretation of brain function, actually sacrifices the main principle of this interpretation. *Any* hypothesis which fits the facts in this field will have the same implication. For, figural after-effects are established only when somewhere in the tissue the level of activity *varies* from one place to another as it does at a contour or an outline. It follows that, apart from events within individual neurons and chains of neurons, differences in a transverse direction must have some specific effects. In this respect our example seems to us so clear that we refrain from quoting other facts which lead to the same conclusion.

There have been two attempts to explain transverse relational determination in nervous tissue. They agree about the main point, the nature of the process which arises when conditions in one part of the tissue differ from those in an adjacent part. In both it is assumed that this process is an electric current. They differ merely in their assumptions concerning the *origin* of this current.

We have felt tempted to avoid any detailed discussion of this problem, which to some degree must remain speculative. For, what we need

in the following is only *some* electromotive force which can drive a current across the tissue when the level of activity in adjacent parts is not the same. Now, it is an old biological rule that local activation of tissue establishes an electromotive force at the boundary between the activated region and its unaffected environment. We may regard it as a highly plausible generalization that any *difference* as to the level of activity in two adjacent areas has the same effect. If the rule and its generalization were accepted no further argument would be necessary. We could simply state that between two adjacent regions of the visual brain there arises an electromotive force if in those regions the level of "brightness processes" is not the same.

With our present knowledge of the nervous system, however, such a short reference to general biological experience will hardly carry full conviction. In recent years we have learned so much about the function of individual cells that any statement about processes in the nervous system must indicate how these processes are related to the activity of the cells as such. It is for this reason that the following paragraphs are written. For our own purpose they have no decisive importance. In the next chapter a theory of visual figures, of satiation, and of figural after-effects will be suggested. In this theory it will be assumed that any inhomogeneity in a visual field is associated with an electric current in the visual brain. The theory will be subjected to a number of tests. But these tests will be concerned exclusively with the assumption as it has just been formulated, i. e., with the existence of such currents, with their distribution, and with their after-effects. As to their origin no more will be assumed—and verified—than that they are caused by brightness differences in adjacent regions. Consequently our theory is independent of any more specific physiological hypotheses with regard to this point. And what we want to show by the following discussion is merely that the theory *can* be connected with available evidence about cell function. We take it as a matter of course that in the future other such connections may prove more valuable than the possibilities of which we are now aware. We repeat that any correction in this respect would have little influence upon our theory.

In our previous report (17; *cf.* also 16, pp. 213 ff.) we have pointed out that so far as chemical action is concerned the effects of impulses in adjacent nerve fibers do not remain separate. When a fiber is activated its surface becomes permeable to various substances which then pass from the fiber into the surrounding tissue fluid. A particular secretion of this kind, that of neurohumors, has recently been regarded as responsible for the transmission of nerve action at synapses. While this particular view has not been generally accepted the fact that during action certain chemicals are liberated by a fiber seems no longer a matter of debate. Suppose now that in a number of adjacent fibers activity is at high level, in other words, that in most of these fibers the frequency of impulses is great. Under these conditions summation of chemical output will lead to a high concentration of secreted substances in the fluid in which these active fibers lie. If everywhere around the active group impulses are less frequent, and travel along fewer fibers, the concentration of secreted chemicals will here be lower. The fact that in both regions these chemicals will be destroyed or removed in several ways does not essentially alter the situation. It merely means that both in the more and in the less active region equilibrium concentrations will rapidly be established. Thus the fluid in the more active region will be characterized by a certain high concentration of such substances, the fluid in the less active region by a lower concentration. Both regions will be in immediate contact at their common boundary in the fluid.

It is a fundamental theorem of physical chemistry that at such a boundary electric forces must arise if the chemicals in the fluid are dissociated into ions. We assume that this condition obtains in the present case. Therefore the difference as to the level of nerve activity on both sides of the boundary will cause the appearance of an electromotive force. The direction of this force is transverse; it points across the boundary of the two regions. Its origin is relational. For its cause is the fact that levels of activity in adjacent parts of the tissue are not the same.

By this force alone, however, no current can be maintained. Any closed line which we draw through the electrically active boundary passes this boundary twice in opposite directions. Under these circumstances there can be no current. A closed flow only becomes possible if at least one more medium is in contact with the two solutions. This implies the existence of further electromotive forces which are located at the

boundaries between the solutions and such a third medium. Now, cell surfaces constitute a third medium with which the tissue fluid both in the more active and in the less active region is in contact. For reasons of differential permeability for various ions (perhaps also because of differential solubility) cell surfaces are known to develop relatively strong electromotive forces in contact with surrounding solutions. The intensity of these forces depends again upon the concentration with which ions are present in the solution. If the two chemically different regions of the tissue fluid, together with adjacent cells, are regarded as a system, closed currents can be established in this system. At first we may find it difficult to visualize just how these currents will actually flow. It will help us to overcome this difficulty if we now consider a second hypothesis about the origin of electric currents in the tissue.

This hypothesis is contained in several papers by R. W. Gerard and B. Libet (5, 6, 19). On the face of it their derivation of direct currents as a major activity of the nervous system may appear entirely different from the one which has just been indicated. It is not so much psychological as strictly physiological evidence by which they are led to their hypothesis. Moreover, while the former hypothesis is meant to refer first of all to the brain of the highest vertebrates, particularly to the visual sector of the human brain, Gerard and Libet develop their ideas in connection with experimental work on the frog's brain. And yet, basically both assumptions may have more in common than a first impression seems to suggest.

To Gerard and Libet transverse interrelation of function within the brain appears stronger and more immediate than can be accounted for by nerve impulses which travel in transverse fibers, and are in this direction transmitted across synapses. Besides, they find that such an interrelation persists when transmission across synapses is prevented by chemical agents or, more radically, by cuts which sever all connecting neurons. Under these conditions ordinary electric currents are probably the only processes by which transverse interrelation can be mediated with sufficient speed. Such currents would pass a cut because when the fibers are severed the tissue fluid still constitutes a continuous conductor.

But what is the origin of these currents? The authors believe that in the neurons of the frog's brain we have to distinguish between two different electromotive forces. One is the force which upon stimulation drives the current of the nerve impulse through the fiber and its environment. As this force changes its location, the nerve impulse travels along the fiber. Quite apart from this, a second electromotive force is postulated which has its seat in the soma of each neuron. More concretely, it is assumed that normal metabolism keeps the dendritic and the axonic poles of the soma at different potentials. Such a force would drive a current through a cell into its environment, and back into the cell, if the environment offered no obstacle to the flow. But in the frog's brain the cell bodies form a sheet, several cells thick, within which all somas have the same orientation. This sheet, therefore, is comparable to a set of galvanic elements, mainly in parallel arrangement, which are so densely placed that almost any line of flow which emerges from one element can return to it only through another element. On this path it would be opposed by the electromotive force of this other element, and if both elements have about the same force, such a circuit would contain two equal electromotive forces in opposition to each other. For this reason no current of appreciable intensity can be established in the brain of the frog until in some region the somas of the sheet lose, or change, their electromotive forces. As soon, however, as in some part of the sheet the cell elements are discharged, or their forces lowered, a circuit is open through which the other unaffected somas of the sheet can drive a current. This current will pass through the unaffected somas in the direction of their axis; it will then flow in a transverse direction toward that "opening" in the sheet, bend and pass through the opening, and return to its source on the other side of the sheet. The current is of course relationally determined. It flows as it does because in two parts of the tissue electromotive forces are not the same.

Under what circumstances is the electromotive force of a soma lowered or lost? It seems plausible to assume that among the causes of such a temporary affection there will be nerve impulses which arrive in the sheet of somas.

This hypothesis has the great advantage of using ideas with which physiological theory is familiar. The authors compare the sheet of brain cells to the surface membrane of a nerve fiber. Just as in this membrane there is an

electromotive force which cannot operate when the force across the membrane is the same all along the fiber, so the sheet of brain cells will remain virtually inactive when the force across the sheet has everywhere the same amount and direction. And again, just as a nerve impulse is immediately established when in some part of the cell surface the potential is altered, so a current can flow through and around the sheet of brain cells as soon as somewhere in this sheet the electromotive forces are affected.

The authors support their assumption by experimental facts. They found that between the dendritic and the axonic side of the sheet of brain cells there was, on the average, a potential of two millivolts. Moreover, when for ten seconds a polarizing current was passed from one side of the sheet to the other its effect depended upon the direction of the current. Although in both directions polarization as such attained the same value, its after-effect subsided much more quickly when the direction of polarization was opposed to that of the normal potential of the cells.

We shall now briefly compare our own hypothesis (p. 324 f.) with the assumptions made by Gerard and Libet. In these assumptions no reference is made to electromotive forces which originate in the tissue fluid when a given group of fibers is more strongly active than adjacent ones. The authors suppose that, quite apart from any action in fibers, the sources of molar currents are always present, and that these sources merely do not operate until a circuit opens. On the other hand, since this hypothesis and the one with which our theory starts are entirely independent and refer to different facts, they may be combined. In the circuit as described by Gerard and Libet we should then have one electromotive force at right angles to a layer of cells, and a second such force, parallel to this layer, which is caused by the fact that on one side of the boundary fiber activity is more intense than on the other. The current would be maintained by the algebraic sum of these forces. In this way the two hypotheses can perhaps be united.[20]

[20] A certain difficulty must be briefly mentioned. If we understand the authors correctly their somatic potentials are short-lived, i.e., after a short time of actual flow these sources are supposed to be exhausted. One more cause for caution is, of course, the fact that findings which refer to the frog's brain are not necessarily applicable to the brain of man.

At this point it will be remembered that a transverse electromotive force which has its seat in the tissue fluid cannot alone establish a current (*cf.* p. 324). We had to assume that the fluid on both sides of this boundary develops further electromotive forces in its contact with cell surfaces. This suggests that the physiologists' hypothesis and our own assumption may have more in common than is at first apparent. One feels tempted to identify those additional electromotive forces of our hypothesis with the electromotive forces postulated by the physiologists. Gerard and Libet would probably grant that their somatic potential is not a matter solely of the soma but of the soma in its relation to the surrounding fluid. The soma potential would therefore vary when the chemical state of the tissue fluid is altered. In this essential respect, therefore, the somatic potential resembles the electromotive force between the tissue fluid and a cell, which is postulated in our hypothesis. From our point of view, on the other hand, such electromotive forces at cell surfaces must exist even when no particular chemical activity affects the surrounding fluid. Consequently, we do not hesitate to admit that, like the somatic potentials of Gerard and Libet, such electromotive forces must have a certain intensity even when all fibers in the neighborhood are inactive. Their intensity would then depend upon the "normal" chemical state of the fluid.

The present discussion leads to a fairly simple picture concerning the flow of molar currents in the tissue. In Figure 67 a sheet of cells is indicated by small circles. Dendrites and axons are omitted. These cells are the seats of electro-

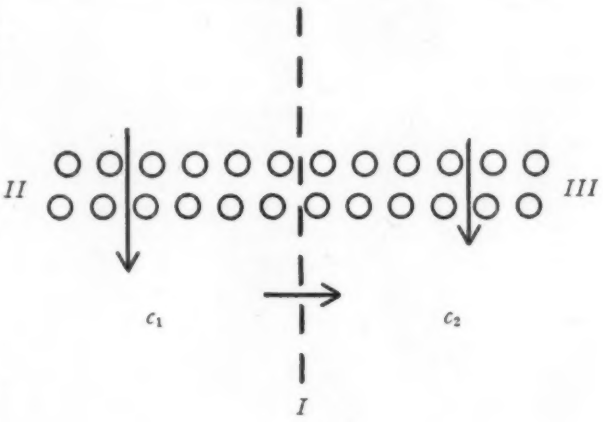


FIG. 67

motive forces even when no nerve impulses arrive in the sheet. But under these circumstances no circuit is available through which these forces can drive a current. Unequal nerve activity on the left and on the right side of a boundary—the broken line of Figure 67—changes this situation entirely. Unequal concentrations, c_1 and c_2, of certain ions are established on one side and the other, and thus an electromotive force (I in Fig. 67) across the boundary arises.

At the same time the electromotive forces of the cells are affected because these forces also depend upon the concentration of ions in the tissue fluid. And since on the left and on the right side this concentration is not the same, those forces (*II* and *III* in Fig. 67) will also differ. Thus a current can be maintained by the forces *I*, *II*, and *III*. This current will have its highest density along the shortest lines of flow, in which the resistance of the circuit is smallest. Obviously, these are the lines which keep near the boundary. Such lines will be surrounded by others which form wider curves with greater resistances. In Figure 68 the places in which

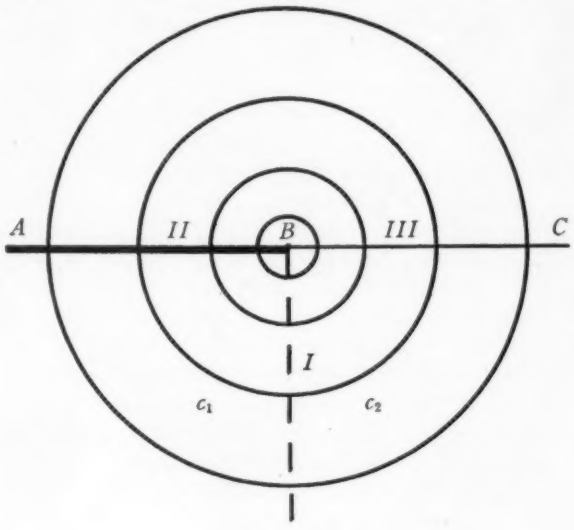


FIG. 68

electromotive forces are situated are indicated by two lines at right angles. The circles in this figure represent the lines of flow, which of course need not actually be circles. The fact that the density of the flow decreases as these circles grow larger is symbolized by an increase of the distances between successive circles.

It should perhaps be explicitly mentioned that in this scheme the tissue is regarded as a continuous volume conductor. Specific resistances may vary from one point to the next, and cell surfaces are invariably subject to immediate polarization. But all parts of the tissue have a certain amount of conductivity. Therefore any currents which spread in this medium must be volume currents.

In the next chapter much use will be made of this simple picture of electric flow in the tissue. We are inclined to believe that the location of the forces which maintain the current is approximately the one which Figure 68 indicates. On the other hand, we wish once more to emphasize that in our further discussions particular hypotheses about the nature of these forces will play no part.

CHAPTER SIX

A THEORY OF FIGURAL AFTER-EFFECTS

I

In Chapter Five material has been presented from which a theory of figure processes, of satiation, and of figural after-effects can be derived. In the present section we propose to take preliminary steps toward such a theory.

1. The functional scheme which was indicated in Figure 67 and Figure 68 allows of immediate application to visual processes. When nerve impulses arrive in the visual cortex they alter chemical conditions in the tissue fluid which surrounds the cells. Suppose that in one retinal area the intensity of stimulation differs from that in another, and adjacent, area. Under these circumstances more nerve impulses will travel in fibers which issue from the more strongly stimulated area. Now, those adjacent retinal areas are connected with adjacent cortical areas. Since more impulses arrive in one of these than in the other, the levels of chemical activity in the two areas will differ correspondingly. Thus we obtain the situation to which Figure 67 refers. As a consequence, a current will flow from one area to the other and back into the first (*cf.* Fig. 68). The flow will circle around the boundary at which the two areas are in contact. Near the boundary it will be dense because here the length of the circuit is short. At a greater distance, i. e., along longer circuits, its intensity will be smaller. In Figure 68, which gives only a cross-section of the situation, the two areas are represented by the lines *AB* and *BC*, their common boundary by point *B*. Actually, the cortical areas are supposed to extend at right angles to the plane of the Figure so that their boundary also extends in the same direction. Accordingly, the flow surrounds the boundary in cylinders rather than in lines. We need hardly mention that even within its plane Figure 68 oversimplifies conditions: the lines *AB* and *BC* represent several *layers* of cells (*cf.* Fig. 67).

The *visual* situation to which this scheme refers consists, of course, of two areas of different

brightness which are in contact along a certain line. The scheme means that, physiologically speaking, any such boundary between areas of different brightness is surrounded by a direct current. At present we have no reason to discuss instances in which areas differ in hue rather than in brightness.

If one of the two areas is *enclosed* in the other it will be seen as a particular object. Figure 69

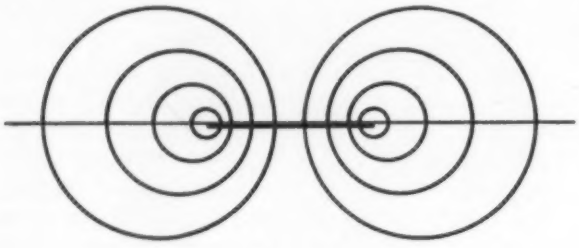


FIG. 69

indicates as a cross-section how the current must flow in this situation. It must circle *everywhere* around the contour of the object. As a consequence the lines of flow can no longer be concentric. Within the limited space of the figure through which the whole flow must pass they will keep nearer the contour than they do in the environment in which they can freely spread. Therefore, the current will also be denser within the object than in the environment.

Our investigation has been principally concerned with outline figures. The current of a single straight line results if in Figure 69 the area of the solid object is reduced to a minimum. We thus obtain Figure 70. With a *closed* outline

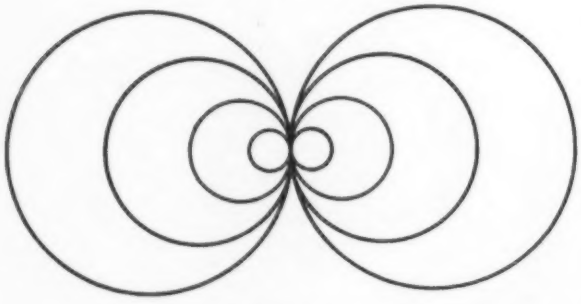


FIG. 70

figure the current will still turn in both directions at each point of the outline; in other words, one part of the flow will pass through the enclosed area and another part through the environment of the figure. But within the enclosed area, which is pervaded by current from all directions, the flow will be denser than it is outside at comparable distances.

It follows from this reasoning that all visual figures are associated with figure currents—provided that the brightness of their area or of their outline differs from that of the ground. In the theory of vision it, therefore, does not suffice to consider cortical color processes at each point. If our assumption is correct we have to attribute equal significance to those currents which pervade the area of visual objects and which also surround them like fields.

Just like all other direct currents, figure currents are functional units: their flow maintains itself relationally in a certain distribution. As a matter of principle this distribution can be mathematically determined. For each figure it is an integral of a partial differential equation. But for the time being an excursion into mathematical physics would be of little help. For, we shall presently see that any figure current must affect its conducting medium, and that as a result its own distribution will begin to change. To follow such a process in strict mathematical reasoning is a task which, to our knowledge, nobody has ever undertaken. Nor will it be necessary to make such an attempt for our purposes. At some points, it is true, precise information about the distribution of figure currents under the conditions of our experiments would facilitate the comparison of observation and theory. But on the whole we shall have to refer solely to the more obvious characteristics of currents in volume conductors.

2. We need no special hypothesis for an explanation of *satiation*. The figure currents of our theory differ in no way from other electric currents. They will therefore polarize all cell surfaces through which they pass and will thus establish the condition which the physiologists call electrotonus. This condition is known to survive its cause, the polarizing current. After intense polarization it persists for considerable periods and affects all currents which then happen to pervade the same region. In the further development of our theory we will assume that what we have so far called satiation *is* the electrotonic effect of figure currents. On this ground, inspection of any specific visual object must cause an electrotonic state within the area of the object and in its environment.

Electrotonus is a heavy word, and no verb has been derived from it. This is most incon-

venient. Although no such difficulty exists in the case of the term polarization we hesitate to use this word as a synonym for electrotonus because the two terms have not quite the same meaning (*cf.* p. 321 f.). In the following pages we shall therefore continue to speak of satiation, with the understanding that from now on electrotonus and satiation will be equivalent terms.

Electrotonus, we remember, is a name for two different facts: an initial, and almost immediately completed, polarization of cell surfaces and a more gradual change of polarizability which is, of course, accompanied by corresponding changes of actual polarization. Clearly, the facts of satiation with which we have so far been occupied must belong to the latter category. Initial polarization develops so fast that if it has any overt effects in human vision such effects would have to be counted among the normal ingredients of visual objects.

As the electrotonic condition develops its spatial distribution in the tissue follows that of the satiating figure current. In places in which the current is particularly dense, electrotonus will develop more quickly and to a higher level than in other parts. Therefore, with solid figures the electrotonic state must reach its greatest intensity about the contour, with outline figures within and near the outline. For the same reason the interior of the figure will be more strongly affected than its environment. At a great distance from specific objects practically no electrotonus will be established because here the figure currents will be too weak.

It would, however, be an ambiguous statement if we were to say that the distribution of electrotonus in the tissue agrees with that of the current by which it was first established. In this statement the current would be regarded as an independent cause and the electrotonic condition as its effect. Actually, since in the electrotonic state the tissue no longer offers the same conditions for the spread of the current as it did in the beginning, the current itself will now be affected by the electrotonus which it has just caused. In other words, under the influence of satiation the distribution of the current will presently change (*cf.* above p. 319). When this happens its further electrotonic effects will of course also be shifted, and so forth. Thus, while we inspect a figure its current can never be strictly stationary. It will always "glide" from an initial distribution to slightly different ones, and if inspection is continued for a long time the change may become conspicuous. Our discussions in Chapter Five allow us to predict in what direction the change must occur. Under the influence of its own electrotonic effects a current will be deflected into regions which so far have been less affected. In this fashion satiation must tend to assume a more even distribution than corresponds to the original distribution of the current. We can easily recognize that this behavior of figure currents and of satiation must have important consequences. The foremost among these consists in the fact that the electrotonic action of currents will always limit itself. For, the more intense satiation becomes in a given place the more current will be withdrawn from this place so that here further satiation proceeds more and more slowly.

The fact that any current in the tissue is bound to weaken itself by electrotonic action will have to be given full attention when attempts are made to register such currents. If figure currents actually exist it cannot be very difficult to demonstrate their occurrence and thus to verify our principal assumption. Both direct-current amplifiers and extremely sensitive voltmeters have been constructed which ought to make such tests feasible. The records, however, could hardly be just like those of ordinary direct currents. Rather, upon the appearance of a figure in the visual field, a kind of on-effect may be expected because as a consequence of electrotonus the first full intensity of the current could be maintained only for a short time. After this initial stage the flow would be reduced to lower intensities.[21]

The description of electrotonic action which has been given in these paragraphs may not be complete. Judging from physiological evidence it seems quite possible that electrotonus not only affects given currents but also establishes currents of its own. In our discussions this possibility will not be pursued but in a further development of the theory it may become necessary to do so.

A much more remote possibility should at least be mentioned because it might be regarded as the starting point for a rival theory of satiation. For reasons which nerve physiologists will fully appreciate, we look upon the following argument with the greatest reserve. According to our theory satiation is caused by figure currents which

[21] We once tried to establish electrotonus in the visual cortex from the outside. Two points of the skull—above one occipital lobe—were connected with the poles of a direct-current source. When, with one subject, fairly strong currents were used the visual field seemed to be affected on the opposite side. But we give no weight to this observation because it ought not to be repeated. For two weeks afterwards the subject suffered from a persistent headache and a disturbing state of depression.

are not to be identified with the migrating currents of nerve impulses. But could not these migrating currents have similar effects? It will be remembered that in the theory of Gerard and Libet (p. 325) each soma of a cortical layer is a galvanic element from which, however, hardly any current can issue so long as the whole dense layer consists of similar elements. Only when at some place the cortical elements lose their potentials will an appreciable flow originate; and then this flow will circle around the boundary between the intact layer and that particular place. Now, at least in a formal way the same situation exists when more or less synchronized impulses travel in a number of parallel nerve fibers. At a given moment a certain stretch in each fiber behaves like a galvanic element. The corresponding current in its temporary location is the electric phase of the nerve impulse. But, if the impulses of all other fibers in the neighborhood have just reached about the same level, the whole region will consist of galvanic elements in a parallel arrangement. As a result only a weak flow will be possible through and around each fiber, and this weak flow, the impulse, will be propagated as the location of the potentials shifts along the fibers. Suppose, however, that as a consequence of differential stimulation in the periphery some among the parallel fibers are not active at a given level when impulses pass this level in other fibers. If we still apply the views of Gerard and Libet, and if we have enough confidence in the nature of nerve impulses *qua* currents, we might say that now these currents can spread more freely and assume a higher intensity. For from each active element the flow of the impulse will extend into the inactive region and then return to its element. Thus, the current as a whole will circle around the boundary between the two regions; it will in this form travel in the direction of the fibers; and again in this form it will arrive in a ganglionic layer. Upon arrival, however, it will have an electrotonic effect upon this layer and, as impulses follow one another at a considerable rate, such effects will accumulate until a fair degree of electrotonus is established. The distribution of the current will determine the pattern of its electrotonic action. Obviously, this pattern will more or less agree with the one which we have derived from our different premise. It is this possibility which we had in mind when we admitted (on p. 323) that satiation within and around specific entities could perhaps be derived from the action of nerve impulses. We hasten to add that this alternative explanation appears to us extraordinarily bold. Against all accepted notions it assumes that the electric flow of a nerve impulse depends upon conditions at a considerable distance from its fiber. Also, the explanation seems to be at variance with the All or None Law. At least in terms of actual flow—as distinguished from voltage—the intensity of a nerve impulse would vary whenever the activity in neighboring fibers changes. Thus nerve impulses would be relationally determined in a fashion for which there seems to be no experimental evidence. Incidentally, if this hypothesis should ever be defended it would not actually contradict our own theory. For, logically, both assumptions could be correct; i. e., satiation could be established in one way *and* the other.

3. Once the electrotonic state has been thoroughly established by a figure current, any further flow in the same region will no longer have the characteristics which it would otherwise have. Owing to the persistency of electrotonus these alterations of subsequent currents must be demonstrable for some time. It is our explanation of figural after-effects that, when inspection of an *I*-object has caused electrotonus of sufficient strength, the currents of *T*-figures are so affected, and that the after-effects are symptoms of this fact. At present we propose to use these effects for a few tests to which our theory of figure currents and of satiation ought to be subjected. No actual theory of figural after-effects will be needed for this limited purpose. It will suffice to assume that they are caused by electrotonus.

In the first place, we have just come to the conclusion that electrotonic action of figure currents will always limit itself (*cf.* p. 329). According to our theory, therefore, if figural after-effects are caused by that action they will appear after comparatively short periods of inspection; they will then grow more conspicuous at a slower rate; and they will eventually reach a maximum beyond which they cannot be enhanced by any amount of further inspection. This conclusion agrees not only with all our own observations in this field but also with certain experiments of Gibson who tested one of his effects after varying periods of inspection and found precisely the development which has just been described (*cf.* 9, p. 459 ff; *cf.* also above, p. 299).

Secondly, the degree of all electrotonic action depends upon the intensity of the polarizing current. This current, on the other hand, is proportional to the electromotive forces which maintain the flow. And the electromotive forces of figure currents grow with the brightness difference between the figure and its ground. It follows that, other conditions being the same, a figural after-effect must be stronger if the brightness difference between the *I*-figure and its ground is increased. This inference was verified in the following differential experiment. On both sides of the fixation mark, on a white screen, outline oblongs of identical size and shape were shown as *I*-objects. The outline of one was black while that of the other was drawn in a light grey which, from a distance of three yards, could still be easily distinguished from the white ground. The *T*-objects were four equal outline squares in the usual arrangement (*cf.* Fig. 14 on p. 279 where, however, only one oblong is shown). After an inspection period

of four minutes the two pairs of *T*-objects were found to differ as expected. On the side of the black oblong the distance of the squares was clearly shorter, and the squares themselves were smaller. The same differential effect was observed when we used *solid I*-figures the brightness levels of which corresponded to those of the outline oblongs in our first observation. A curious phenomenon, which the *I*-patterns of these experiments exhibit at first sight, will be discussed in the last chapter.

Thirdly, it can make no essential difference whether the *I*-figure is shown dark on a bright ground or bright on a dark ground. The electromotive forces must be the same although in the second case their direction must be reversed. This merely reverses the direction of the current; but apart from that the spatial distribution of the flow will remain about the same. Its electrotonic action will therefore also occur in the same distribution, and as a result similar after-effects must be observed in the two situations. We here presuppose that in both cases the brightness relation between figure and ground is the same in the *T*-figure as it is in the *I*-object. Our prediction agrees with Gibson's observations. It has also been verified in numerous experiments of our own. White objects on a dark ground give figural after-effects of the same kind as dark objects on a white ground. As a single exception we may mention that, while with dark outline figures on a white ground the dark *T*-object tends to look too grey, i. e., too bright, with white lines on a dark ground the "color effect" is not that the white *T*-lines are darker. Rather, in a satiated area, these lines appear "weakened" or too thin.

We may now turn to a fourth point to which we attribute particular significance for our theory. In Chapter Five we mentioned that after prolonged flow of a current in the tissue the electrotonic condition at anodes, the anelectrotonus, becomes more and more predominant. At this stage anelectrotonus means an obstruction for all test currents, irrespective of their direction. Now, if figural after-effects are caused by electrotonus they must be caused mainly by anelectrotonus in this advanced stage (*cf.* p. 329). It follows that such figural after-effects will be observed whether the direction of the *T*-currents is that of the *I*-currents or the opposite. But, as we said in the preceding paragraph, the direction of a figure current will be reversed if the brightness relation between the figure and the ground is reversed. Thus we are led to conclude that if, for instance, an *I*-figure is given black on white and the corresponding *T*-figure white on black, we must obtain the same figural after-effect as though the *T*-figure were also black on white. Actually, Gibson mentions observations in which this was the case. In our own experience after-effects were under these circumstances about as strong as they usually are. Only the color symptom must again be mentioned as an exception: when, after inspection of a black outline figure on a white ground a white outline figure on a black ground served as a *T*-object, the white lines of this object were never affected. Their color was the same as that of neutral comparison lines. This observation is particularly interesting because it shows once more that other symptoms of satiation are independent of the color effect. We wish to emphasize that this holds also for the "depth effect" which remains as conspicuous under the present conditions as it is when brightness relations in the *I*-pattern and in the *T*-pattern are the same.

We can go one step farther. Physiological investigation has shown that even alternating currents cause the electrotonic condition so long as the number of the cycles is not too high. Direct test currents are afterwards affected more or less as though a direct current of a certain intensity had previously passed through the tissue. The obvious inference is that if *during the inspection period* the direction of a figure current is often reversed we must still obtain about the same after-effects as though the direction of that current had been constant. Now, the figure current assumes the opposite direction whenever we reverse the brightness relation between the *I*-figure and its ground. According to our theory figural after-effects must therefore still be observed if during the inspection period a dark *I*-figure on a bright ground alternates with an identical bright *I*-figure on a dark ground.

It will suffice if we mention only a few experiments in which this conclusion was tested. In a grey screen two equal openings of oblong shape were cut. The right opening was permanently filled with the black of a second screen behind the first. Behind the left opening black and white surfaces appeared in regular alternation. The *T*-pattern again consisted of four squares having the position as indicated in Figure 14 (p. 279) for one oblong. When with a total in-

spection time of five minutes the brightness relation on the left side was reversed every two seconds, no difference between the *T*-squares on the left and the right side could afterwards be discovered. Even with several reversals per second the outcome of the experiment remained the same. Since we know that the after-effect in the area of the permanently black oblong was strong, it follows that in spite of the reversals the after-effect was equally strong on the other side.

It is a *general* rule that after-effects are not prevented if during the inspection period the brightness relation between the *I*-figure and its ground is frequently reversed. In another test the *I*-trapezoid of Figure 55 (p. 306) was projected upon a screen so that the figure appeared alternately black on white and white on black. When during a total inspection time of twenty seconds reversals had followed one another at a rate of one per second, the objectively parallel *T*-lines clearly converged toward the wider side of the *I*-figure.

II

Our experiments have shown that the various symptoms of satiation, the color effect, the depth effect, and displacements do not depend upon one another. We are, therefore, justified in now selecting for theoretical treatment one of those phenomena and in postponing the discussion of the others. To the color- and depth-effects we have so far given scant attention. Thus our theoretical discussion of figural after-effects will be mainly concerned with the *displacements* which *T*-objects suffer in affected areas.

All our observations of displacements refer to phenomenal facts. On the other hand, our assumptions about figure processes and the nature of satiation refer to brain events, which are physical facts. Our observations and our theory can, therefore, be related only by means of a psychophysical axiom. This axiom must indicate what event in the brain we assume to correspond to a displacement in the visual field, and vice versa. Without such a principle we could never decide whether observed displacements agree or disagree with the consequences of our theory. Plainly, in choosing a cortical counterpart of visual displacements we take a step the implications of which reach beyond the present investigation of figural after-effects. There can be only one brain correlate of visual displacements. It must be the same whether a displacement is caused by satiation or in any other way whatsoever. We, therefore, propose to consider in the next paragraphs the correlate of visual displacement in general.

Displacements of visual objects are changes of their position. But if an object changes its position it does so in reference to other objects. Thus our psychophysical axiom must state what physical fact in the brain we intend to correlate with the relative position of visual objects. Once this correlate is chosen the psychophysical representation of *changes* of relative position, i. e., of displacements, will follow directly.

Spatial relations are of two classes. In a first class belong relations like "inside," "outside," "adjacent," and so forth, which are called *topological* relations. Topological relations are characterized by the fact that they are not affected by distortions of the metrical characteristics of space. In the second class belong all *quantitative* and *semi-quantitative* relations in space such as relations of distance, area, and volume. If we say that a given distance is three times as great as another distance we refer to a quantitative relation. In the statement that a given distance is greater than another distance the relationship is semi-quantitative. In the following discussion we shall deal with such semi-quantitative relationships. But we shall include the particular case of *equality* which may be regarded as belonging to either variety.

Nobody seems to doubt that the neural correlates of such visual facts as hue and brightness are phases of actual physical processes in the brain. We assume that the same holds for the correlates of *all* aspects of visual experience, including experienced spatial relationships. Elsewhere (16, p. 220) this rule has been applied in a discussion of topological relations. The correlates of such relations, the rule says, cannot be mere geometrical facts in the brain; rather, they must be *functional* characteristics of brain processes just as the correlates of hue and brightness are. Take as an example two visual objects which have separate positions in the visual field so that neither they nor their retinal images touch each other. At present it is generally recognized that in the cortical projection of the retina topological relations are preserved. With only one exception adjacent portions of the retina are represented by adjacent parts of the visual cortex.[22] Conversely, if two retinal

[22] The exception consists in the discontinuity of the visual cortex in the median vertical.

images have separate positions the same holds for the corresponding processes in the brain. Under these circumstances objects which have separate locations in the visual field will also be represented by separate processes in the cortex; i. e., some other processes will be placed geometrically between them. According to our rule, however, this geometrical fact as such has no psychophysical significance because it does not constitute a *functional* reality.

But geometrical separation of two cortical processes may have a functional counterpart. In spite of the fact that such processes have separate locations, they may be functionally interrelated. This is possible if the presence of each leads to a field action in their common environment, by which a functional interrelation is mediated. On the other hand, if the processes in question have separate locations in the tissue they will be functionally interrelated *solely* by their field action.

There are, however, *degrees* of cortical separation; or, if we prefer a positive formulation, processes in the visual brain occur more or less close to each other. For the sake of brevity local cortical processes which correspond to specific visual objects will from now on be called "cortical objects." Cortical objects, then, are neighbors in varying degrees. Again, this is not a fact to which we attribute direct psychophysical significance. It is a much more important fact that *if cortical objects approach each other in the brain, their functional interrelation will be strengthened, while with an increase of their distance that interrelation will be weakened.* All physical field action depends in this fashion upon distance.

But what makes the distance between cortical objects vary? For the most part the cause will be that the distance between corresponding retinal images varies. Generally speaking, therefore, if the distance between two retinal images changes, the degree of functional interrelation between the cortical objects in question will also change. For instance, smaller distances between images on the retina are likely to mean more intense interrelations between cortical objects.

There is, however, a further correlation: We know that *visual* distances tend to vary as retinal distances vary. Thus, when the degree of functional interrelation between cortical objects changes because the distance between corresponding retinal images changes, the distance between the visual objects in question will generally vary at the same time. Conversely, and more specifically, when two objects approach each other in visual space functional interrelations between their cortical counterparts are likely to be strengthened.

We will go beyond this statement. We will adopt it as an axiom that *degrees of closeness in visual space are psychophysically represented by degrees of functional interrelation in the visual cortex.* In another formulation: We will assume that *distances between visual objects shrink or grow as functional interrelations between their cortical counterparts are strengthened or weakened.*

It will be remarked that in this formulation distances in the cortex are no longer mentioned. Our axiom refers only to degrees of functional interrelation between cortical objects. To be sure, the degree in which cortical objects are interrelated *may* be changed by a change of their distance. But it may also be changed by other causes. We do not intend to apply our axiom only in the former case.

There is no reason why the scope of the axiom ought to be restricted to the case of separate visual objects. The parts of one extended object also lie more or less close to each other. If we wish to be consistent we shall have to say that under these circumstances functional interrelations among the parts of the cortical object vary just as the closeness of corresponding parts in the visual object varies, and vice versa.

In many instances the consequences of our axiom will not be in the least surprising. Suppose, for example, that a cortical object (1) is placed in a region of the visual cortex which offers in every part and in all directions the same conditions for the spread of interrelating field action. In such a medium, where must we place two further cortical objects (2) and (3) if the functional interrelation between the objects (1) and (2) is to be the same as that between the objects (1) and (3), and therefore the visual distance (1) (2) the same as the visual distance (1) (3)? Evidently, functional and therefore visual symmetry will obtain when the cortical distances (1) (2) and (1) (3) are equal. For, under the conditions of our example, functional interrelations in a homogeneous medium are graded in strict accordance with geometrical distances.

But geometrical distances in the brain and degrees of interrelation by field action will no longer agree if the conditions for the spread of

this action vary from one cortical region to another. Let us assume that between the cortical objects (1) and (2) those conditions are more favorable than they are between the objects (1) and (3). Even if the geometrical distance (1) (2) is exactly the same as the distance (1) (3), functional interrelations will now not be symmetrical. The objects (1) and (2) will be more strongly interrelated, and in the visual field the distance (1) (2) will appear shorter than the distance (1) (3). Conversely, the same degree of functional interrelation will now obtain across two cortical distances which are geometrically *un*equal. For this to happen, either the distance (1) (2) will have to be lengthened, or the distance (1) (3) to be shortened. But in spite of this fact, after the correction the visual distances (1) (2) and (1) (3) will appear equal because now functional symmetry has again been established.

So far our psychophysical axiom has been stated in extremely general terms. Actually, our conclusions in section I of this chapter allow us to give it a more specific formulation. In our theory the field action which pervades and surrounds a cortical object is its figure current. Interrelations among cortical objects are therefore interrelations which are maintained by their currents. This gives us the following version of the axiom: *The distance between visual objects varies with the degree to which corresponding cortical objects are interrelated by their figure currents.*

There is one term in this statement which has not been clearly defined. The precise meaning of functional interrelations has not been stated. And this omission is made particularly conspicuous by contrast. For otherwise our theory uses more concrete concepts than is common in psychophysics. But the omission is not at all dangerous. It has purely practical reasons. The functional interrelations to which we refer are fully recognized in the physical theory of direct currents. In fact, they are identical with the relations by which currents maintain their own distribution. Thus there need be no fear that the term hides any mysterious notions. On the other hand, its meaning cannot be made perfectly explicit without a discussion of fairly technical matters. Such a discussion would carry us too far away from the program of this investigation. It is fortunate that our work will not be impeded by the omission. At no point shall we have any difficulty in deciding whether a given change of conditions enhances or weakens functional interrelations between given cortical objects. And this is the only issue with which our theory will deal in its present stage.[23]

[23] Although the term "functional interrelations" need not at once be more precisely defined, it seems advisable to mention that a further clarification will have to concentrate upon interrelations within one figure current rather than upon any interrelations between separate currents. Our formulations in the text are chosen for the purpose of an initial simplification of thought. It is in this sense that cortical objects are said to be interrelated by *their* figure currents. Actually, even if several figures are seen as discrete entities there are no separate figure currents but only one current, which is a coherent flow in their common ground, and then divides into branches, corresponding to the individual figures. As a consequence, the interrelations among such figures are essentially the same as those among the parts of one figure.

Although our psychophysical axiom refers to functional relations rather than to anatomical data we will now introduce certain assumptions about the medium in which those relations obtain. For indirectly the characteristics of this medium *are* relevant to our theory in that they influence the distribution of figure currents. And upon the distribution of these currents depend the interrelations with which our axiom is concerned.

1. The experimental situations of our observations were not defined in cortical terms. They were defined with reference to patterns which our subjects inspected on screens. It is true that, under the conditions of those experiments, such patterns were, roughly speaking, repeated on the retina by optic projection. In our theory, however, we intend to discuss the behavior of figure currents in the brain. Clearly, in order to refer the theory to retinal conditions, and thus to patterns on screens, we must have a rule which indicates how retinal and cortical areas are quantitatively related. In this respect we now propose to make the simplest possible assumption, an assumption which we know to be incorrect. We shall do so for a good cause. To consider functional interrelations in the brain as the correlates of relative position in visual space is not an activity to which we are accustomed. In the beginning it will require a certain effort. If we were at once to introduce anatomical complications we might not be able to master the unusual task at all. In a temporary fashion we shall therefore assume that, when nerve impulses project retinal areas upon the visual cortex, the metrical proportions of the retinal patterns are always preserved in the projection. In other words, we shall postulate that retinal patterns and their cortical projections are geometrically similar. Later it will be seen that our theory itself tends to eliminate the error which we now deliberately commit.

2. We have to add a second assumption be-

cause the first does not yet enable us to deal with figure currents in a definite relation to retinal data. We know that if the tissue is an inhomogeneous conductor, i.e., if the various parts of the visual cortex do not offer the same conditions for the spread of currents, the distribution of these currents will be distorted. For under these circumstances equal geometrical stretches, areas, and volumes are no longer equivalent conductors. It will again simplify our thought if in this respect also we make the simplest possible assumption. We therefore postulate that in all parts of the visual cortex the conductivity of the tissue is the same. This hypothesis has been tacitly adopted in Figures 68, 69, and 70. It is most probably not correct. In the next chapter it will also be replaced by a more realistic assumption.

When taken together our two assumptions mean that patterns in the visual cortex are, both geometrically and in terms of conductivity, strictly equivalent to the retinal patterns to which they are related. As a consequence figure currents in the cortex will be distributed just as though the cortex coincided with the retina, and as though cortical objects were identical with retinal images. We have to emphasize, however, that one characteristic of cortical tissue is not affected by our idealizing postulates: *microscopically*, the visual brain remains an extraordinarily heterogeneous medium. Any macroscopic fraction of its volume contains a great number of cells, a number which we assume to be everywhere the same; and the surfaces of these cells are subject to the electrotonic action of figure currents. But as soon as these currents establish the electrotonic state in a certain area, our second assumption no longer applies. Rather, the affected area will be temporarily transformed into an inferior conductor, and figure currents will react to this condition. Thus, our second assumption is meant to be valid only so long as subjects do not fixate a point within or near a visual object.

III

With the discussions of section II our preparations have been completed. We can now explain why as a result of satiation *T*-objects tend to be displaced in visual space. Such an explanation will, of course, have to indicate in what particular positions relative to satiated areas displacements of *T*-objects actually occur. In this fashion we must obtain an explanation of the distance paradox and of the fact that displacements are maximal at a certain distance from *I*-objects. Naturally this will lead to a discussion of the rule that in affected areas small angles are increased. Moreover, once we know why there is a distance of maximal displacement we must also be able to explain why with mere lines as *I*-objects the distance of maximal displacement is *greater* than it is with solid figures.

1. We have merely to connect a number of statements which have previously been made in order to explain the displacement of *T*-objects in general and the size effects which result from such displacements in particular. From section I of this chapter we take it as our premise that satiation is the electrotonic condition which the figure current of an *I*-object establishes in the visual cortex. But from the same section we know that so far as our experiments are concerned that condition must affect test currents irrespective of their direction. In this sense, then, there is no practical difference between electrotonus (or satiation) and a state of increased resistance.

In section I of Chapter Five we have, however, seen that, if currents spread in a medium which has not the same resistance throughout, their distribution will be altered. It will differ from the distribution which would obtain in a homogeneous medium. More concretely, while current will be withdrawn from regions which have a higher resistance, the flow will be intensified in others which offer less resistance.

This rule we can now apply to test currents which spread in the visual cortex when in certain parts of the tissue the resistance has been raised by *I*-currents. Quite generally, the total intensity of such currents will, of course, be weakened. But as a rule their distribution will also be *distorted*. The flow will be reduced in strongly affected regions; and it will turn with increased intensity in directions in which the tissue has been less or not at all affected. Whether in a given case such a distortion actually occurs must depend entirely upon the position which the *T*-object has with regard to the affected zone.

Now, in the preceding section of this chapter a correlate of relative position in visual space has been chosen. Distances between visual objects, we have decided, vary with the degree of functional interrelation which figure currents maintain between corresponding cortical objects. But, when the distribution of *T*-currents is

distorted by previous satiation, the interrelation of the cortical *T*-objects in question will be altered. It follows from our psychophysical axiom that under these conditions the *T*-objects themselves must be seen at abnormal distances from each other. But this is tantamount to saying that in such situations displacements of *T*-objects must be observed.

Take the very simplest example: On a screen an object (1) lies half-way between two other objects (2) and (3). The screen, we shall presuppose, is shown right in front of a subject so that the retinal distances (1) (2) and (1) (3) are also equal. According to an assumption which we have made in the preceding section the corresponding cortical objects have now the same symmetrical distribution. But we have also assumed that, in the absence of satiation, the visual cortex is a homogeneous conductor. As a result, the distribution of the figure currents must be just as symmetrical with reference to the object (1) as is the geometrical situation. But then the same must hold for any functional interrelations across the distances (1) (2) and (1) (3). The subject will therefore report that the visual distances (1) (2) and (1) (3) appear equal.

He will no longer say so, however, if he sees the same objective pattern after previous inspection of a figure in the area between the objects (1) and (3). The current of such an *I*-figure raises the resistance in this area. When the objects (1), (2), and (3) are now given as *T*-objects their currents cannot be symmetrically distributed. Less current than before will flow in the region between the cortical objects (1) and (3): the current of object (3) will be deflected away from object (1); and the current of object (1) will turn with increased intensity in the direction of object (2) (*cf.* Fig. 71, in which the withdrawal of flow from the area between the objects (1) and (3) is indicated by the arrows). Interrelations between the objects (1) and (2) will thus be enhanced, while those between the objects (1) and (3) will be weakened. Hence, according to our psychophysical axiom the subject will now report that the distance between the *T*-objects (1) and (2) appears to him shorter than the distance between (1) and (3).

This example contains our explanation of numerous figural after-effects which have been described in previous chapters. Distances between separate *T*-objects may in this fashion be lengthened or shortened; and while in our example both effects are combined, the explanation remains the same if, with other *T*-patterns, the lengthening and the shortening of the distances are separately demonstrated. For instance, our explanation demands that the distance (1) (3) is lengthened in any case, whether or not the object (2) is present. Consequently, if that distance is compared with a *neutral* standard it must appear longer than this standard. Our observations have shown that this is true. Similarly, in the absence of object (3) the distance (1) (2) must also be shortened, particularly if on its other side, beyond object (2), a second *I*-object were shown so that the current of the *T*-object (2) would be deflected to the right.[24]

It will not be necessary to show how our explanation applies in a variety of patterns such as have been used in our experiments. The model given in Figure 71 illustrates the general principle from which, according to the individual situations, different distortions of *T*-patterns can be derived. We shall merely add that this principle works just as well with closed *T*-figures as it does with separate *T*-objects. For instance, in Chapter Two we demonstrated that a *T*-circle which is shown in the interior of a larger *I*-circle appears considerably too small. This must be so since inspection of the larger circle establishes a closed zone of increased resistance around the *T*-circle. Now, the outline of this *T*-circle sends one part of its current into the environment while another part flows through the interior of the circle. When the resistance around the

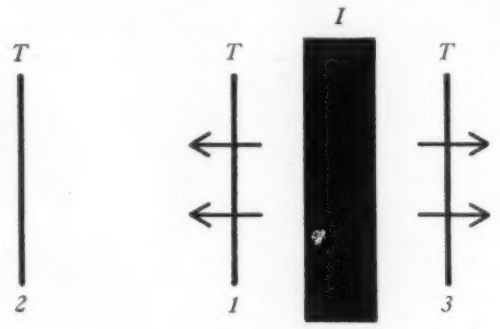


FIG. 71

[24] In such a pattern the two *I*-objects ought not to lie too near each other. If they do, their currents will raise the resistance not only on the left and on the right side of the stretch (1)(2) but to a degree also within this region. It is obvious that under these conditions the distance (1)(2) will not be maximally shortened (*cf.* p. 296).

T-circle has been increased the former part of its flow will be weakened, but the current in the interior will be intensified. This strengthens the interrelations between opposite parts of the circle, and its diameter is accordingly reduced. Just the opposite must happen if the *T*-object surrounds a previously shown *I*-object. The interior current of such a *T*-object will be weakened; thus interrelations between opposite parts of this object will also be weakened; and as a result the object as a whole will be visually enlarged. That this is indeed the case has been shown in Chapter Four (p. 296).

In the present theory we presuppose that any distortion of visual patterns after satiation is caused by a change of functional interrelations among the parts of corresponding cortical patterns. We do not assume that in figural after-effects these parts themselves are displaced in the tissue. Now, it would be premature to maintain that the relation between retinal locations and cortical positions is absolutely rigid. There may be conditions under which that relation proves to be flexible within certain limits. But in this investigation we shall not refer to this possibility because it seems more important to keep one principle quite clear: Even if, with a constant retinal image, cortical projection could be slightly shifted, a corresponding visual pattern would be changed not because of that fact as such, but because this fact would affect functional interrelations among the cortical objects.

A further problem which we do not wish to raise at this point is concerned with the comparison of affected with *neutral* *T*-objects. Suppose that an affected *T*-object appears larger than its neutral partner. This is a relation in visual space; and if our theory is to be accepted it must be capable of being applied not only to sizes as such but also to size relations which are expressed in comparison judgments. We quite approve of this conclusion. It will, however, be understood that on this new ground we must try to restrict our first steps to simpler tasks.

2. We have not yet explicitly stated under what conditions an individual *T*-object will assume a changed position. This is our task now.

Suppose that a single horizontal line is given as an *I*-object, and that a second, and parallel, line is used as a *T*-object. What will happen if this *T*-line is shown in varying positions relative to the *I*-line?

As a first instance, let the distance between the *T*-line and the *I*-line be very great. Under these circumstances the current of the *T*-line will barely extend into the region in which the *I*-current has raised the resistance of the medium. Moreover, the parts of the *T*-current which spread so far would be extremely weak even if no satiation had occurred in that region. Therefore the distribution of the *T*-current will remain virtually unaffected, and the *T*-object itself will appear in its normal position. For its interrelations with neutral comparison objects in its environment will not be altered.

But there is also another, and entirely different, position of the *T*-line in which it cannot be displaced. This position is that of coincidence with the *I*-line. If the *T*-line has this location, resistance is maximal within this line because the *I*-current had its greatest density here; and on both sides of the *T*-line zones of less but still considerable resistance extend symmetrically. Now, this is one of the situations in which questions of physics can be decided by a mere look at given configurations. As the physicists express it: The *T*-line cannot be displaced one way or the other for reasons of symmetry. If a displacement in one direction is tentatively suggested the question at once arises why the displacement does not occur in the opposite direction, in which conditions are exactly the same?[25] In terms of our theory, the current of the *T*-line will be weakened; but it will remain symmetrically distributed so that the interrelations of the line with objectively equidistant objects on both sides also are symmetrical. This makes the corresponding visual distances equal, and the *T*-line will stay in its normal location in visual space.

We may now return to our starting point, the position of the *T*-line at a very great distance, where the line is practically unaffected. Let us imagine that from this position the line is step by step brought closer to the *I*-line. As we proceed, more and more of the *T*-current which would normally flow in the direction of the affected area will now turn away from this area and into less affected zones, particularly in the opposite direction; for, more and more, that current meets with the increased resistance of the satiated area. Any interrelations with neutral objects will be altered accordingly. Especially, relations will be strengthened with figures which lie *beyond* the *T*-line on the same side of the *I*-line. From the point of view of our theory, therefore, the *T*-line will be displaced; it will appear in a location which is too distant from the place of the *I*-line.

We can easily show that this explains not only displacements as such but also the distance paradox. To repeat: A *T*-line will not be displaced at a very great distance; it will also

[25] For physicists we will remark that this position of the *T*-line represents a *stable* equilibrium.

not be displaced when it coincides with the *I*-line; but it will recede from the *I*-line in intermediate positions, in which it lies at moderate distances from the *I*-line. In a short formulation, this *is* the distance paradox, although now there is nothing paradoxical about it. It will, however, be advisable to consider more in detail what happens when the position of the *T*-line approximates coincidence with the *I*-line. So far we have found that the displacement of the *T*-line must first grow as the distance between this line and the *I*-line is gradually shortened. In what way, then, is the displacement again reduced to zero when that distance is further lessened?

The answer is obvious: With decreasing distance from the *I*-line the displacement of the *T*-line will grow so long as one part of the *T*-current, the flow in the direction of the *I*-line, is progressively more affected, while at the same time the flow in the opposite direction still passes through much less affected tissue. But as we bring the *T*-line nearer and nearer, the satiated *I*-region will gradually extend *beyond* the place of the *T*-line. The more this is the case the less will the *T*-current be deflected, and the less therefore will the line be displaced, because the position of the *T*-line within the satiated *I*-region begins to be progressively more symmetrical. This development will continue until the resistance on one side of the *T*-line is just as great as it is on the other side. This happens in the position of coincidence where no displacement can occur.

It will be recognized that according to the present explanation the displacement of the *T*-line must be maximal at a certain distance from the *I*-line. *Two* factors, we have just seen, determine that displacement. On the one hand, if a displacement is to occur the *T*-current must penetrate into the *I*-zone in which the resistance is strongly heightened. Otherwise the *T*-current would not be noticeably affected. To this extent a shorter distance between the *T*-line and the *I*-line will make for greater displacement. On the other hand, the *T*-current only on one side of the *T*-line ought to be strongly affected if the displacement is to be considerable. Therefore the *T*-line must not lie *too* near the *I*-line, since in the immediate neighborhood the *T*-current would meet with about the same obstruction in both directions. Thus, as we approach the *I*-line, the distance of maximal displacement will be passed when the negative effect of the latter factor just begins to overcome any positive effect which further approach might have because of the former factor. Beyond this point, at still shorter distances, the displacement must obviously decrease.

3. We can now discuss the fact that a *T*-line which forms a small angle with an *I*-line always turns away from this line. Essentially, the explanation of this phenomenon follows from our explanation of the distance paradox. The *T*-line cannot be displaced at the point at which it intersects the *I*-line because here the obstruction is the same on both sides of the *T*-line. But the more we proceed to the free end of the *T*-line the more the zone of heightened resistance lies on one side of this line. Consequently, its current will be increasingly deflected away from the *I*-line so that the various parts of the *T*-line recede by correspondingly growing amounts. This interpretation would hold so long as the distance between the ends of the *I*- and the *T*-line is not too great; in other words, so long as the angle is fairly small, or as the lines are not too long.

Up to this point the argument is hardly open to criticism. But we cannot regard it as a complete account of the factors which make a *T*-line turn away from an *I*-line. When we discussed the experiments which referred to the growth of small angles we remarked that no analysis of the effect into separate displacements at various points of the *T*-line seemed to us in order (*cf.* pp. 286 and 303). More particularly, we said, satiation opposite a point of the *T*-line is not determined by a single point of the *I*-line which may, geometrically, be regarded as opposite. Furthermore, we added, a *T*-line is surely too coherent for independent displacements of its parts to be possible. The former thesis is fully borne out by the explanation of satiation which we have since given. In the case of an actual *I-line* the concentration of the *I*-current opposite a given point of the *T*-line is obviously much higher than it would be if the *I*-line were reduced to a mere point at that place. The same holds, of course, for the intensity of the electrotonic state which is established by that current. We approach a more important theoretical issue if we now consider in what physical sense the *T*-line must be regarded as a coherent entity. The answer is that, just like any other direct current, a *T*-current represents a functional unit which maintains its distribution relationally. If we apply this to its behavior

in the situation now under discussion we come to the following conclusions. The cortical *T*-object is a place through which all flow of the *T*-current must pass. From this place the current spreads into the environment, as it were, in an enormous number of circuits which are arranged in parallel.[26] Now what happens if the resistance in some of these circuits is heightened? The current is partly withdrawn from these particular circuits and it turns with increased intensity into others in which the resistance has remained unchanged. In this statement we seem merely to repeat the characteristic of electric currents which is crucial for our theory as a whole. On closer inspection, however, the statement points to a fact which has been ignored in the argument of the preceding paragraph. There it was tacitly assumed that if a part of the *T*-line has a certain position relative to the satiated region, this condition will affect only the current which issues from that part. But this is against the principle upon which all our explanations rest and which we have just repeated. For, increased resistance in the neighborhood of a given part of the *T*-line has the further effect that, if other parts of the *T*-line lie in a better conducting region, the flow in such other parts will be intensified.

Consider the point at which the *I*-line and the *T*-line intersect. Here the *T*-line cannot be displaced.[27] None the less, the fact that at this point the *T*-current is greatly obstructed contributes to the turning of the line. For, parts of the flow which is here obstructed will now be shifted to other parts of the *T*-line which lie in a less affected region. But we realize that even if such a region is less affected it still offers more resistance on one side of the *T*-line than it does on the other. The current which is "transferred" to this region will therefore mainly flow toward the side on which the resistance is lower, in other words, away from the *I*-line. Clearly, the same must happen all along the *T*-line. Every part of this line which lies in a region of higher obstruction must transfer some of its flow to neighboring parts where the obstruction is lower. And since this obstruction has an asymmetrical distribution with regard to the *T*-line the transferred current will flow more strongly toward the side of the line on which the resistance is lower, i.e., away from the angle. In this fashion all parts of the *T*-line which lie more deeply in the affected zone will contribute to the displacements which occur nearer the free end of the line.

The turning of a *T*-line, therefore, does not allow of quantitative comparison with the displacements which a *T*-line suffers in the neighborhood of a *parallel* *I*-line. It is, for instance, possible that a *T*-line turns *more* than could be expected if the displacements at various parts of the line were mutually independent. And a distance which gives the maximal displacement when the *T*-line is parallel to the *I*-line need not be a distance of maximal displacement when the two lines form an angle. This may be the explanation of the fact that in our observations about the growth of acute angles the free end of the *T*-line was never less displaced than were other parts which lay nearer the *I*-line.[28]

As to the particular form which the turning of a *T*-line assumes near an *I-curve*, our argument in Chapter Four (p. 307) remains valid. If with a straight *I*-line a straight *T*-line remains approximately straight as it turns, then with a curved *I*-line the *T*-line *must* curve in the opposite direction.

4. As a last peculiarity of displacements we shall now discuss the finding that the distance of maximal displacement is much greater if, with the *I*- and the *T*-object in parallel arrangement, the *I*-object is a line rather than a solid figure. The explanation of this fact follows from the way in which the figure currents are distributed in the two situations.

Take a solid oblong as an *I*-object. We remember that in the case of a solid figure the lines of flow circle in every direction around the boundary of the figure (*cf.* Fig. 69, p. 328). With an oblong of very great width they are free to penetrate deeply into this area. In this respect a change will occur if we reduce the width of the oblong. The lines of flow surround not only one boundary but also the boundary on the opposite side. Within a narrower oblong, their spread around one boundary is therefore limited by the fact that the lines of flow which surround the opposite boundary claim half of the interior for *their* spread. As a consequence the lines of flow in a given direction will no

[26] We use this expression merely for those who are not accustomed to thinking about coherent flow in continuous volume conductors.

[27] If there is any displacement it cannot be in the direction in which we are now interested.

[28] Partly, also, the reason may be that in our experiments the free ends of the *I*- and the *T*-line lay in the periphery of the field where the distance of maximal displacement is always greater (*cf.* p. 303).

longer have their former position with regard to the boundary. Within the oblong they will be "crowded" in the sense that they are here forced to keep nearer the boundary than they would in a wider oblong. But this must have an effect upon those parts of the lines which pass through the environment of the oblong. Suppose that the boundary which we are considering is kept stationary. Then, as the width of the oblong decreases the lines of flow must move *toward* the boundary in the *interior* of the figure, but they will move *away* from the boundary in the *environment*. In other words, the pattern of the flow will be pushed toward the outside when the oblong shrinks. Eventually, the solid oblong will be reduced to a line. In this case the pattern of the flow will be farther displaced in the same direction than it is with any oblong. But, when the pattern of the current moves, the distribution of its electrotonic action must move in the same direction. Now, we have seen that the displacement of a *T*-object depends upon its position with regard to the affected area. If this area is shifted in a certain direction, given phases in the displacement of the *T*-object will occur at places which are shifted in the same fashion. Therefore, when a solid *I*-object shrinks, the distance of maximal displacement must grow, and with an *I-line* the distance of maximal displacement will reach its greatest size.

It will be realized that in this explanation we do not discuss the absolute *amounts* by which the *T*-objects are displaced. With narrow oblongs, and particularly with a line, the total intensity of the current may be lower. For this reason its electrotonic action may also be weaker; and as a result, *T*-objects may on the whole be less displaced. There is, of course, no contradiction between this statement and the thesis that the distance of maximal displacement is now greater (*cf.* p. 302).[29]

Does the present theory of figural after-effects give any hope of being applicable to the color- and the depth-effect? We shall limit our answer to a few remarks.

A figure current connects a figure and its environment. Its immediate action in the tissue is electrolytical in the sense that the concentration of certain ions is heightened or lessened at interfaces. This, we know, gives rise to polarization and electrotonus. But we must not forget that when ionic concentrations vary at a given point, chemical reactions at this point are likely to be affected by the fact. If some chemical reactions in the tissue are the psychophysical correlates of colors we must, therefore, expect that colors in the visual field are to a degree influenced by the existence of figure currents. The intensity of these currents, on the other hand, depends upon the brightness difference between a figure and its environment. Therefore, if achromatic color processes are at all affected by figure currents, this influence will grow when that brightness difference grows. This seems to give us a theoretical access to achromatic color contrast—especially since contrast is known to depend upon specific configuration, as it must if figure currents are its cause.

So far, it will be realized, these statements refer to *immediate* effects of figure currents. They are not directly related to any after-effects. But figure currents gradually satiate their own medium and thereby weaken their own intensity. Therefore, after prolonged satiation, contrast ought also to be weakened inasmuch as it is a function of those currents. Now, the color black is known as a typical contrast effect. It follows that in a satiated area black should degenerate and become grey. We know that this is actually the case, and that no explanation in retinal terms is acceptable. For, this color effect of satiation is observed even if the eye which is used in the test has been excluded from inspection of the *I*-object (*cf.* p. 277).

An explanation of the depth effect will not be possible until we are able to indicate quite generally how the third dimension of visual space is psychophysically represented. In this respect we can only say that any figure current in the tissue has three dimensions. Not only does it spread in a medium which extends in three geometrical dimensions (we know that this fact can be related to visual space only in an indirect fashion), but a figure current has also three dimensions in a *functional* sense. For, the flow at a given point occurs functionally on the same level with the flow at certain other

[29] In section I of this chapter we pointed out that during prolonged inspection of a figure its current is likely to shift from its original distribution to different ones because any existing distribution of the current is soon affected by its own electrotonic action. Thus the distribution of satiation must also gradually change. In the explanations of this section we have not referred to this fact. It was not necessary to do so inasmuch as with *I*-oblongs or mere *I*-lines the change consists mainly in a simple widening of the satiated area.

points if these points happen to lie on the same equipotential surface as the first point. It does not matter whether the surface in question is geometrically a plane. Now, while in this fashion only two functional dimensions of the current are given, its third dimension in a functional sense obviously coincides with the direction in which the gradients of potential lie; in other words, in which the actual flow occurs. If this description of figure currents can be related to the three dimensions of visual space then the depth effect must be caused by the fact that satiation influences the flow of test currents in all their three dimensions.

CHAPTER SEVEN

THE THEORY CONTINUED

Our theory cannot be left in the form which we have given it in Chapter Six. It necessitates further conclusions which we shall now discuss. But it will also be remembered that as an explanation of figural after-effects the theory rests on certain idealizing assumptions. These must now be eliminated.

I

On various occasions we pointed to the persistency of satiation. Under favorable circumstances this condition was found to survive from one day to the next—in fact, sometimes longer. In the present section we propose to defend a much more general statement. *A certain degree of satiation*, it seems to us, *belongs to the normal and permanent make-up of the visual cortex.*

Specific visual objects of one kind or another are in our visual field from early in the morning until the end of the day. Every thing, figure, spot, or line establishes its figure current as soon as it appears; and each figure current has an immediate electrotonic effect upon the tissue, not only at the place of the object, but also in its environment. It does not matter that cortical objects are seldom stationary because we rarely fixate a given point for any length of time. To be sure, when a cortical object changes its position in the visual brain its current and the electrotonic action of this current will move with it. But as we turn our eyes, such cortical objects as have just left a given cortical region will generally be replaced by others which continue the satiating action of their predecessors. For this to happen no actual coincidence with those predecessors is needed since the figure current of each cortical object satiates a sizable area. For the same reason, even if a single object moves across an otherwise uniform region, given cortical points in and near its path will be subject to satiation during a certain period.

Homogeneity is not a frequent condition in the visual field. Our average environments being what they are, it is a safe conclusion from what we have just said that, so long as we keep our eyes open, no part of the visual cortex will for long be entirely exempt from electrotonic action. As a result there must be a general accumulation of satiation all over the cortical field. It would be inconsistent to admit that under experimental conditions a stationary cortical object satiates the tissue and yet to deny that under conditions of everyday life the whole visual cortex is to a degree kept satiated by a virtually continual flow of figure currents.

This argument is supported by good empirical evidence. There can be no doubt that electrotonic action accompanies the very onset of currents, and that upon repeated exposure to short-lived currents higher degrees of electrotonus are built up by summation. As to the particular case of satiation in vision, it will be shown in section III of this chapter that very short exposures of figures cause some degree of satiation, and that repetition of such exposures rapidly enhances this condition by accumulation.

It cannot be objected that in the course of a day currents of opposite direction will pass through a given part of the visual cortex with equal frequency, intensity, and duration. Owing to the asymmetry of electrotonic action the effects of currents which successively flow in one direction and then the other do not cancel each other. In the long run the anelectrotonic state must everywhere prevail.

As a name for the electrotonic state of the tissue which is thus maintained by normal visual processes we suggest *statistical* or *permanent* satiation, because this state is bound to be permanent in an extreme sense. To our knowledge, no other factors are so favorable for persistency of satiation as continual or repeated operation of its causes. It would be at odds with all our experiences in this field if, when a whole day's figure currents have acted upon the visual cortex, the resulting satiation were to disappear overnight. And if it does not so disappear, the next day's figure currents will rapidly compensate for any loss which may have

occurred during sleep. On the whole, therefore, permanent satiation will be kept at a fairly constant level, fluctuating moderately as sleep interrupts vision or as unusual environments may occasionally raise or lower the average intensity of figure currents.

We may now turn to the *distribution* of permanent satiation. To begin with a simple case, we shall first presuppose that our two idealizing assumptions are valid, namely, that cortical patterns are geometrically similar replicas of retinal patterns, and that the conductivity of the visual cortex is everywhere the same.

What, statistically, is the distribution of retinal patterns in the course of a day or during a longer period? There is no major regularity in the distribution of objects which confront human beings. We can therefore take it as a first approximation that, if a sufficient period is considered, images are evenly scattered on the retina. If this premise is combined with our assumptions about the visual cortex it follows that, in a statistical sense, cortical objects, their contours, outlines, etc., are also evenly distributed in the brain. As a result the same must hold for the distribution of figure currents and for that of their electrotonic action. In other words, under such circumstances permanent satiation will affect all parts of the visual cortex to the same degree. We are still supposing that the cortex is a homogeneous conductor in its original state. According to the present argument it will remain such a conductor in the state of permanent satiation. The only thing that will now be changed is the *degree* of cortical conductivity. In all parts of the visual cortex the resistance of the tissue will be equally raised when permanent satiation has been established. Under its influence all further figure currents will, of course, be weakened; but their distribution as such will not be affected. It would be affected only if resistances were asymmetrically altered in the environment of given cortical objects, Since they are not so altered all patterns of interrelation will remain the same, and therefore no distortions of figures are to be expected.[30]

Can we make any statement about the *level* at which visual processes keep the tissue permanently satiated? In order to answer this question we must first consider whether the concept of permanent satiation interferes with our theory of figural after-effects. It will be apparent that so far this theory is not in the least affected. In developing the theory we made no assumption about the conductivity of the tissue excepting that it was supposed to be the same at all points. If it remains the same at all points when the tissue is permanently satiated, our theory obviously applies just as though permanent satiation were absent. We must only remark that from our present point of view any satiation which is established in an experiment on figural after-effects represents a mere *increment* that is locally added to the level of permanent satiation, while in a "neutral" region this level is likely to be more or less preserved during the inspection period. Once this is clear, the explanations which we have given in the preceding chapter need not be modified. Satiation *can* be locally raised beyond its permanent level because inspection of a circumscribed figure on a homogeneous ground, with fixation constant, does not occur in natural vision. As an exceptional and optimal condition for satiation it *must* cause an increment of satiation—unless we maintain that in its permanent form satiation has reached its highest possible level.

We see no reason to defend this radical thesis. On the other hand, the level of permanent satiation ought not to be underrated. We know that the electrotonic action of figure currents tends to limit itself. In actual tests (*cf.* 9, p. 459 ff.) figural after-effects were found to appear after fairly short inspection periods, and then to become more conspicuous only with marked deceleration. If the concept of permanent satiation is sound, such data refer to an increase of satiation from a given permanent level to an exceptional temporary level which is established under special circumstances. It follows that at a lower level, but under otherwise comparable conditions, satiation must develop much faster than it does with permanent satiation as a starting point. Also, in the absence of permanent satiation figure currents must be stronger, and their electrotonic action more powerful, than is possible when, with permanent satiation established, the self-limiting behavior of the currents is more fully operative. The inference appears plausible that the level of permanent satiation is comparable to any increments which we can locally add to it in our experiments. And it may be that the level of permanent satiation is much *higher*.

So far we know only what effect permanent

[30] There is a possibility that the *size* of visual objects will be reduced to a smaller scale.

satiation would have upon an idealized visual cortex which is largely of our own construction. Actually patterns in the cortex are not geometrically similar to retinal patterns; and we doubt whether, with or without satiation, the cortical tissue has everywhere the same conductivity. It will be our next task to show that, to the degree to which permanent satiation affects the conductivity of the tissue, those simplifying assumptions are now no longer needed. For, an important functional characteristic of permanent satiation makes them partly or wholly superfluous. It can be proved that, if cortical patterns are geometrically distorted pictures of retinal patterns, permanent satiation must invariably tend to compensate for the distortions. Furthermore, if the conductivity in certain parts of the visual cortex is not originally the same as that in other parts, then the difference will always be reduced by a counteracting distribution of permanent satiation. What we propose to demonstrate is, therefore, that with regard to the relations between the retina and the visual cortex *permanent satiation operates like a homeostatic device* in Professor Cannon's sense. But it seems to us a matter of some importance to biology in general that in actual fact there are no devices which force electrotonus to work in that direction. If we are right, permanent satiation must have such regulating effects upon the visual cortex for purely functional reasons.

Let us separate the two parts of our task and discard at first only the assumption that in its original state the conductivity of the visual cortex is everywhere the same. For the moment we shall still assume that, geometrically, cortical patterns are similar projections of retinal patterns. Then the question arises how in that original state, i. e., apart from permanent satiation, vision is affected by local variations in the conductivity of the tissue. We know the answer from previous discussions (pp. 333 f. and 336). Suppose a circle with a point in its center is presented on a uniform ground and parallel to the face of a subject. Since we still assume that cortical projections are similar to retinal patterns, not only the retinal image of the figure but also the corresponding cortical object will be circles. But the current of this object will be distributed in accordance with the conductivity which the various parts of the medium possess. For instance, even though the geometrical distance between the center and the circular outline is constant in all directions, the functional interrelation between the center and that outline will vary if the conductivity of the tissue is not also the same in all directions. Functionally, the outline will be farther away from the center where the conductivity of the intervening medium is lower. For, some current will be withdrawn from such places, and the flow will be intensified in directions in which the conductivity is higher. Under these circumstances, the visual figure cannot exhibit the symmetry which characterizes the physical object, its retinal image, and the cortical object. In what particular distortion it will actually be seen depends, of course, upon the particular pattern of local conductivities at the place of the cortical object and in its environment. Similarly, under the same conditions of presentation the objective middle of a straight line would not as a rule appear as its visual middle, although the corresponding point in the cortical object would be the geometrical middle of this object.

All this, we repeat, holds so long as the visual cortex is an inhomogeneous conductor in which permanent satiation has not yet developed. But it *must* develop, and we can now indicate what its effect upon the cortex, upon figure currents, and therefore upon vision will be. First of all, since we are still assuming that there are no geometrical discrepancies between retinal and cortical patterns, any rule as to the distribution of retinal images is *ipso facto* also a rule as to the distribution of cortical objects. If, as we presuppose, retinal images are in the course of time evenly scattered on the retina, the density with which cortical objects occur in the various parts of the visual brain must be just as uniform. To this extent then there is no cause for an uneven distribution of permanent satiation by the figure currents. Nevertheless, there will be local variations in the degree of permanent satiation because the intensity of figure currents, and therefore of their satiating effects, depends upon the resistance with which those currents meet at one place and another. In places in which the resistance is higher the currents will be weaker, and will affect the tissue less, while in places of lower resistance both the currents and their satiating action will be stronger. This holds also statistically because the frequency with which figure currents of all kinds and intensities occur in the various parts of the cortex is throughout the same. It follows that any differences as to the local conductivity of the tissue must be gradually compensated for as

permanent satiation develops. For, this factor will raise the resistance in given parts of the tissue more or less in inverse proportion to the initial resistance in those parts. Theoretically, the process will continue until the tissue has become a uniform conductor. At this stage satiation will everywhere proceed at the same, and of course lowered, rate. The theoretical ideal will actually be reached if, originally, local conductivities have not varied too much or, conversely, if the final level of permanent satiation is sufficiently high to compensate for any differences which originally existed. At any rate, about the *direction* in which permanent satiation must operate there can be no doubt: if in the beginning the visual cortex is not a homogeneous conductor, permanent satiation will always tend to transform it into a homogeneous medium. Visually, this means that such distorting effects of uneven conductivity as have been described in the preceding paragraph will gradually disappear when permanent satiation develops, because the asymmetries of functional interrelation which are responsible for those distortions will be more and more abolished. In case the final stage of the process is actually reached *no* such distortions will be left, and visual figures on a homogeneous ground will share the symmetries of corresponding objective patterns. Moreover, if later any temporary influence should again disturb this optimal condition, the statistical effects of satiation will invariably correct the disturbance. Thus, the behavior of permanent satiation is truly homeostatic.

In passing we may remark that for the same reason the statistical effects of satiation must be counted among the factors which obliterate given figural after-effects. Such effects occur when locally a pattern of heightened satiation has been established. But when fixation is discontinued all further figure currents will be weakened in the affected area, and they will turn with increased intensity into other regions. Statistically speaking, therefore, satiation will be relatively increased in these regions, and it will be relatively weakened in the zone of the *I*-figure. Thus the difference between this zone and the other regions must gradually be lessened. The same holds, of course, for any differences of satiation which inspection has established among parts of the affected area itself. They must gradually be obliterated by the tendency of further figure currents to pass through the less affected parts.

We now know that permanent satiation must tend to compensate for any local variations which the conductivity of cortical tissue may originally exhibit. In the following we shall therefore presuppose that the visual brain is a homogeneous conductor. Naturally, our next question is how, statistically, satiation reacts upon *geometrical* discrepancies between retinal and cortical patterns. The answer will be facilitated if we first consider how such discrepancies must affect vision originally, i. e., in the absence of permanent satiation. As an example we may again take a circle which is presented at right angles to the median plane so that its retinal image is also a circle. But we no longer assume that the corresponding cortical object has the geometrical symmetry of a circle. Now, any distortion of the cortical object can be described in terms of distances. For instance, geometrical distances which are equal in a circle will generally not be equal in the distorted cortical object. But in a homogeneous conductor field action, such as the spread of figure currents, depends upon distance, and so does any interrelation which is maintained by their action. Consequently, functional interrelations among the parts of the distorted cortical object cannot exhibit the symmetry which would make the visual figure a circle. Let us consider the various diameters of the figure. Since they have not generally the same length, functional interrelations along them cannot be the same. Thus, the object cannot be seen as a circle. In the other example which we previously used there will also be distortions: the objective middle of a straight line will not appear as its visual middle if the two halves of the line have not the same cortical length. Nor can we expect the line to appear as straight. With a metrically irregular projection of retinal distances upon the visual brain the cortical line is likely to be a curve. And if the medium is a homogeneous conductor, functional interrelations within the figure current of a curve cannot have the symmetry which would make the visual object a straight line.

In spite of all this the circle on a homogeneous ground is, of course, seen as a circle; the objective middle of a single straight line does appear approximately as its visual middle; and if the line is objectively straight it does appear as practically straight in the visual field. It has been said that any field theory of vision will meet with the greatest difficulties when it begins to face such facts. Actually, however, the difficulties are no greater for a field theory than they are for any other theory of vision. Fundamental symmetries of simple objective patterns are, under the conditions to which we now refer, amazingly well preserved in vision. And yet in

the geometry of cortical patterns this correspondence is surely for the most part not preserved. This is a problem which *all* theories have to solve. We will now show that permanent satiation must tend to compensate for the distortions of visual objects, which the distortions of their cortical counterparts would otherwise cause.

In the theory of figure currents field action depends upon distance only to the extent to which distance is a measure of resistance. And distance measures resistance solely in a homogeneous conductor. We have just seen (p. 343) that in an inhomogeneous conductor interrelations across equal distances are not equally intense. Conversely, in such a conductor interrelations are the same across certain unequal distances. To return to a former example (p. 336), if in the visual brain the conductivity between an object (1) and an object (2) is greater than it is between the object (1) and an object (3) which lies on the other side of (1), then functional interrelations across the distances (1) (2) and (1) (3) will be equally strong when these distances are made unequal. The distance (1) (2) will have to be lengthened or the distance (1) (3) shortened until the two distances are equivalent with regard to interrelating field action; in other words, until the resistance across the distance (1) (2) is the same as that across the distance (1) (3). When this has been achieved the two distances will be visually alike. In this fashion differences as to conductivity can be compensated for by differences as to distance. But this implies that *differences as to distance can* also *be compensated for by differences as to conductivity*. In fact, our example proves the latter possibility as well as the former.

With this in mind we will now discuss the distribution of permanent satiation when cortical objects are geometrically distorted versions of retinal images in a homogeneous conductor. We again start with the premise that no position of an image on the retina is *a priori* more probable than any other position; in other words, that in the course of time the retinal distribution of contours is statistically uniform. Our emphasis upon contours is proper since satiation by figure currents is most intense in the neighborhood of the contours.

It will immediately be recognized that in a visual cortex which renders retinal patterns in a distorted fashion satiation *cannot* be statistically uniform. Those distortions are *systematic* in the sense that *all* cortical patterns which occur in the cortex are distorted according to a common principle. The principle is given by the scale in which any retinal distances are represented in given parts of the cortex. We can apply this to the distances between contours. And then we realize that, if on the retina the incidence of contours is statistically uniform, it cannot be uniform in the cortex. In parts of the cortex which render given portions of the retina in a small scale the average density of the incidence of contours must be higher than it is in parts which render those retinal portions in a large scale. Or also, the distances between both simultaneously and successively given contours will, statistically, be smaller in cortical regions which represent related retinal portions in a small scale. As a consequence the spatial density of figure currents must, statistically, be greater in such small scale regions, more or less in the proportion in which the scale is here smaller. But then their satiating action must also be statistically stronger in the same regions; and if so, the electrotonic resistance of the tissue will here begin to rise above the resistance in large scale regions in which the density of figure currents is lower.

The meaning of this in visual terms will at once be apparent if we again return to the example in which, at right angles to the median plane, the points (1), (2), (3) delimit two objectively equal distances (1) (2) and (1) (3). Visually these distances are originally not equal if, in a homogeneously conducting tissue, the cortical distance (1) (2) is shorter; because under these conditions the interrelation by currents is stronger across the distance (1) (2) so that this distance is also visually shorter. But we have just concluded that inasmuch as the tissue in the direction of object (2) is obviously a small scale region, cortical distances in this region must be *statistically* shorter, and that, therefore, figure currents on this side of the object (1) must *on the average* be denser. Since it follows that, statistically, satiation affects this side more strongly than the other side, the functional interrelation between such objects as (1) and (2) will gradually be weakened. Thus, while originally the distance (1) (2) was visually shorter than the distance (1) (3), the difference would be found to decrease if the same test pattern were given at various times during the devlopment of permanent satiation.

We have first seen that permanent satiation

must tend to abolish differences of cortical conductivity, and therefore visual distortions, when the scale of cortical projection is uniform. We can now add that in a homogeneous tissue permanent satiation must tend to compensate for local variations in the scale of cortical projection, and therefore to abolish visual distortions which would otherwise be caused by this factor. Actually, of course, both regulating effects of permanent satiation must occur at the same time.

The homeostatic effect with regard to differences in the scale of cortical projection is achieved by an increase of resistance in all areas in which the scale of projection is smaller than it is in their neighborhood. This development will gradually approach, but need not generally reach, a stage in which the distortions of cortical objects are just compensated for by counteracting differences in the local level of permanent satiation. If that stage is actually reached visual patterns on a homogeneous ground (and parallel to a subject's face) will show the same symmetries as characterize corresponding objective patterns and their retinal images. Now, minor deviations from the ideal are empirically known. The present theory has good reasons for admitting their existence because our argument has simplified conditions in certain details. It is not quite correct to say that the incidence of retinal objects is statistically uniform. Rubin has rightly pointed out that in the case of man lower parts of the environment tend to contain more particular objects than higher parts (22, p. 83). Moreover, when objects are projected upon the periphery of the retina their images have not quite the same size as they have nearer the fovea. Again, although just as many objects may be projected upon the periphery as upon foveal and parafoveal regions, the periphery of the eye and these central regions are not functionally equivalent. We have no right to assume that the figure currents of given objects are equally effective when their contours are dim. In such respects the theory remains subject to minor corrections. However, for the moment it will suffice if merely its main principles are considered.

For the purpose of future tests a few consequences of the theory of permanent satiation should be briefly mentioned. It would be difficult to test this theory in the normal visual field of adults because permanent satiation is supposed to be effective mainly in making adult vision normal in the sense in which it appears to us normal. But there are a few situations in which, as a matter of principle, an examination would be feasible. In the *first* place, the theory suggests that very early in life the visual field of a child is not yet normal if adult vision is used as a standard. Soon after birth, when permanent satiation is not completed, objects may show distortions which will later disappear. It is to be regretted that at this early period adequate observations may be *practically* impossible. *Secondly*, such distortions might be discovered if a blind-born person is operated on at an age at which he can serve as a subject. This presupposes, of course, that before the operation his lenses have been sufficiently opaque to prevent virtually any formation of contours on the retina. If under such conditions distortions were actually found to exist, the way in which they disappear would be just as relevant to the theory as their original existence. Incidentally, if our theory is correct nothing could be more natural than that after the operation such people do not recognize visual forms in their relation to previous tactual and kinesthetic experiences with the same physical objects. Such a lack of recognition is far from proving that an empiristic interpretation of visual space is justified. As a *third* possibility it will be remembered that in certain instances patients are not allowed to use their eyes for weeks on end. We do not know to what degree permanent satiation survives such a situation. If it is markedly lowered, cases of this kind would afterwards offer a chance of discovering whether with a reduced satiation of the whole field symmetrical objects exhibit any distortions which are normally absent.

In concluding we will remark that according to the present theory the famous alternative between a nativistic and an empiristic explanation of visual facts must be regarded as seriously misleading. If we are right about the effects of permanent satiation, visual objects may originally be quite different from what they are at a later stage. None the less the change would not be attributable to anything like learning processes in the usual sense. We should also hesitate to regard the regulating influence of permanent satiation as a case of maturation. For, if in psychology the term maturation refers to late stages in the ontogenetic development of the tissue, then permanent satiation can hardly be said to be an instance of maturation. We may add that statistical satiation is not a process for which a particular hereditary equipment of the species can be made responsible. Original discrepancies between retinal and cortical patterns, for instance, are sure to differ from one individual to another. No rigid inherited device of a given species could deal with such individually varying situations. Permanent satiation, on the other hand, will exert its homeostatic influence in a manner which fits the particular conditions of each individual case.

So far the greatest discrepancy between the distribution of images on the retina and their cortical representation has not been mentioned. We have not referred to the fact that the visual cortex consists of two areas which are separated by the median fissure. It is well known that vision fails to exhibit any obvious symptoms of this anatomical situation. Figures with cortical

counterparts lying partly in one and partly in the other hemisphere are just as continuous as are others with cortical representation not so divided; also, in such instances, patterns preserve given objective symmetries; and grouping of discrete objects occurs entirely undisturbed by the same factor. We by no means hold that we are able to explain this paradox.

The problem cannot be solved by a reference to the *corpus callosum* the fibers of which connect the two hemispheres. If a colored area extends from one half of the visual field into the other the continuity of its color shows no interruption in the region of the median vertical. Now, it will generally be assumed that in man psychophysical color processes occur only in cortical tissue, and that no such processes are possible in structures like the *corpus callosum*. Therefore, any functional connection which may be mediated by this part of the brain will not help us to understand the continuity of colored areas the cortical representation of which is interrupted by the median fissure. Are we wrong in assuming that in man color processes are restricted to cortical tissue?

In another respect, it is true, our theory does modify the theoretical problem which is offered by the separation of the two striate areas. At first sight it may appear as though functional interrelation across the median vertical must necessarily be less intense than interrelations within one hemisphere. But we are not convinced that this view is correct. Permanent satiation is bound to keep the conductivity of cortical tissue at a comparatively low level. On the other hand, if any currents flow parallel to the fibers of the *corpus callosum*, from one hemisphere to the other, their electrotonic effect within this structure will be small because such currents do not pass through highly polarizable cell bodies. In other words, the conductivity for currents which pass from one striate area to the other is likely to be considerably *greater* than that with which the currents meet in cortical tissue. In a functional sense, therefore, the *corpus callosum* is likely to constitute a particularly short connection. Moreover, figure currents which issue from cortical objects in the neighborhood of the median vertical will preferably turn into this structure with its high conductivity. Such considerations may serve to make the intimacy of figural relations across the median vertical less paradoxical than it is generally believed to be.

Even so the situation obviously remains enigmatic. We do not feel that a satisfactory explanation will be possible until two points have been clarified. First, we shall have to know whether, as many neurologists believe, the foveal regions of the two eyes actually have a representation in both striate areas. It is obvious that the answer to this question would be relevant to the solution of our problem. Secondly, much will depend upon the results of further investigations with patients whose *corpus callosum* has been cut as a surgical measure (*cf.* Smith and Akelaitis, 25). If their visual fields should continue to yield no evidence of any disturbance, radical changes in our views about the neurology of human vision may become unavoidable. In the meantime any further speculation about this problem would seem to us premature.

From an empirical point of view we have to ask ourselves whether the effects of visual satiation extend from one hemisphere into the other. The answer is clearly that they do. We know, for instance, that black outline figures which are shown in affected areas appear pale in comparison with *T*-objects in a neutral zone. The same effect is observed if the *I*-figure lies near the median vertical in one half of the field while the crucial *T*-object is given in the other half but also near the median vertical. In fact, under these circumstances the effect is often particularly striking. At the same time the distance between the fixation mark and the two objects may be sufficiently great to exclude, with trained observers, an explanation by lack of accurate fixation. Equally striking is the depth effect. *T*-objects recede in the third dimension if they lie in the neighborhood of a previously shown *I*-figure. They also do so, and often strongly, when the median vertical separates the loci of the *T*- and the *I*-objects. Lastly, from experiments which have been described in Chapter Five it can be concluded that under the same conditions *T*-objects recede from the area of *I*-objects. After satiation small angles which are bisected by the median vertical grow about as much as do the same angles in a horizontal orientation (*cf.* pp. 310 f.). Such displacement effects across the median vertical are not restricted to instances in which the *T*- and the *I*-object have at least one point in common. Suppose that an *I*-oblong is shown in one half of the field, and that its longer edges are parallel to the median vertical. The satiating effect of this figure displaces a properly

located *T*-square which is shown in the other half of the field. Compared with a neutral *T*-object such a square lies too far away from the *I*-object. But for unknown reasons the effect is clear only if the distance between the *I*- and the *T*-object is considerable.

II

We may now return to *temporary* satiation which, according to the preceding section, is locally added to permanent satiation whenever we inspect a particular object. So far we have mainly been concerned with the figural after-effects which are caused by temporary satiation while the alterations which given figures suffer under the influence of their own satiating action have been given less attention. It will be remembered that Gibson uses in this connection the term adaptation. But for reasons which will presently be stated we prefer to speak of distortions by *self-satiation*.

From a theoretical point of view the effects of self-satiation involve no new principles. When figures change themselves by their own satiating action their distortions follow the same general rules as have been found to hold for figural after-effects. If we, nevertheless, propose now to consider self-satiation *per se* we do so first of all for a practical purpose: it seems that self-satiation is an unrecognized but necessary by-product of many experimental techniques which are being used in the investigation of visual problems.

Surely, temporary satiation is not restricted to experiments in which we aim at the demonstration of figural after-effects. The only condition which must be fulfilled if temporary satiation is to be established consists in fixation of a point when one or several objects are given in a visual field. But we prescribe fixation in numerous experiments. In fact, careful control of fixation is often regarded as an elementary demand of exactness. And in certain investigations the very sense of the problem makes such a control a necessary measure. Unfortunately, if fixation is thus assured, a certain amount of temporary satiation will invariably follow.

Against this one cannot argue that for the most part subjects are not asked to keep fixation constant for a long while. We know that an inspection period of five seconds may suffice to cause a figural after-effect. In such a short period, therefore, temporary satiation reaches a considerable level. Moreover, quantitative experimentation in perception as a rule involves a statistical technique. It is customary to have a subject make several observations with a given pattern in a sequence; and this makes accumulation of satiation unavoidable even if each individual period of fixation is fairly short. It is, therefore, not in the least doubtful that self-satiation accompanies certain standard procedures of experimentation in perception.

In many instances this fact may do no harm. Some visual functions which are tested with those procedures may not be essentially affected by satiation. Others may be influenced only if satiation reaches degrees which will rarely occur in common experimentation. But at the present time we do not know just when we may ignore the presence of that fact. Obviously, investigations which refer in any way to visual brightness do not belong to the "safe" class. For it can readily be shown that after a period of fixation the brightness of an object tends to be affected not only by "retinal fatigue" but also by satiation. In view of this fact the results of few experiments with prescribed fixation can be regarded as being strictly above suspicion.

The situation is not improved by the fact that for the purpose of measurement certain parts of a given experimental pattern are usually varied within limits. This excludes neither satiation nor its accumulation because in each case the satiating action spreads so far into the environment of lines and contours that individual effects overlap. Any subsequent presentation of the pattern in question will therefore occur in an area of badly defined *compound* satiation, and in this sense the subject's judgment will now depend not only upon present conditions but also upon a number of others as represented by the accumulated satiation effects of previous presentations. Thus, instead of being improved, the situation is now merely less clearly defined than it would be with a strictly constant experimental pattern.

These remarks, we hope, will not be interpreted as expressing a vague negativistic attitude with regard to experimentation in perception. There can be no negativism because for the most part it will not be particularly difficult to change experimental techniques so as to meet the danger once it is clearly recognized. And if there is criticism of customary procedures it is surely not vague. Quite probably several phenomena which now puzzle psychologists are principally caused by satiation. Their interpretation may

be greatly facilitated if in the future experiments are planned from this point of view. From among the suspected facts we will here consider one instance in which we can immediately show that the interpretation in terms of self-satiation is correct.

It has often been remarked that some of the most familiar "illusions" tend to become less conspicuous if, for the purpose of measurement, the patterns are inspected in a great number of successive trials. This holds, for instance, for the Zöllner-, the Hering-, and the Poggendorff-patterns. But to our knowledge the most astonishing reduction occurs in the case of the Müller-Lyer "illusion" the almost complete disappearance of which under conditions of repeated observation was demonstrated by Judd (*cf.* 30, p. 647). Up to the present time this phenomenon is being regarded as a practice effect, a most unfortunate interpretation which has done much to support the view that those "illusions" represent errors of the observer rather than genuine visual facts. Truly visual facts, it seems sometimes to be thought, are as hard as diamonds and cannot easily be altered when conditions of stimulation are constant. Errors of observation, on the other hand, can be overcome by practice. But so can the distortions which we call "illusions." They are, therefore, caused by psychological processes rather than by sensory dynamics.[31]

At the present moment we do not propose to explain the Müller-Lyer illusion. This illusion we will take for granted. What we wish to explain is merely its gradual reduction in the course of continued or repeated observation.

In quantitative experimentation the two crucial distances of the Müller-Lyer pattern are usually given as parts of a straight line (*cf.* Fig. 13, p. 278). Whether the line is actually drawn, or whether the distances are left empty, is immaterial for our present purposes. Obviously, in this configuration the figure current must be more concentrated in the area which is to a degree surrounded by oblique lines. Therefore, satiation must also be stronger on this side; and in Chapter One we have shown that it actually is. During one short period of observation the difference may have little effect. But if one observation follows another as it does during measurements, the fact of consistently asymmetrical self-satiation must cause a distortion. All *T*-objects recede from more strongly affected areas into less affected zones, because their figure currents turn in the latter direction. Now, the arrow-heads on the more closed side may be regarded as such *T*-objects. Consequently, these parts of the figure will gradually shift away from each other. But this means that the distance which first appears shorter will gradually be lengthened. Thus the illusion will steadily decrease.

If this interpretation is correct the so-called practice effect must also be established if we simply fixate the middle of the figure for a few minutes. Observation confirms this expectation: after a while the illusion is clearly less striking than it was in the beginning.

It is a further consequence of our explanation that when, after repeated observations, the illusion has been greatly reduced it must again resume its full size if now the pattern is turned by, say, 90° so that its two halves are placed in unaffected—or, partly, equally affected—areas. We need not perform this experiment because it was done by Judd himself who really found this effect, and thus confirmed the present explanation at a time when the concept of satiation was still unknown. It seems possible that the reduction of the illusion could be almost entirely prevented if after each measurement the relative position of the two parts of the figure were reversed. For under these circumstances satiation would on the average affect both parts in about the same fashion.

We may add a further test. From our point of view, precisely what is the situation at the time when, either after many measurements or after prolonged inspection, the illusion has practically disappeared? We maintain that at this time the factors which cause the illusion are still operative, but that their influence is compensated for by the asymmetrical distribution of satiation, which has the opposite effect. Conversely, we may say that at this stage the effect of asymmetrical satiation is approximately balanced by the factors which cause the illusion. It is, however, plainly the presence of the angles which originally causes the illusion. Hence it is also their presence which now balances the effect of satiation. Thus, if after prolonged or repeated observation the angles are omitted, the effect of asymmetrical satiation

[31] It is in defense against this interpretation that in the above paragraph quotation marks have been added to the term illusion. They can from now on be omitted because it must be clear that we are using the old term without the implications with which it is usually associated.

must immediately appear alone, and therefore in full force. In other words, the distance which originally looks shorter in the illusion pattern must now be longer, and the originally longer distance of the illusion must now be shorter. We tested this inference. As a *T*-pattern we used three short vertical lines which passed through the apices of the three angles so that after omission of the angles the two objectively equal distances of the illusion pattern were still delimited. The effect was perfectly clear: After prolonged inspection of the Müller-Lyer figure the two distances in the *T*-pattern differed; and the direction of the difference was the opposite of the one which is observed in the illusion.

We regard these facts as sufficient evidence for our statement that unless particular measures are taken the repetitive schemes of experimentation to which we are accustomed will always introduce the potent factor of self-satiation. Moreover, if the presence of this factor is not recognized we may easily give our findings an entirely mistaken interpretation. We have surely done so when calling the gradual disappearance of the Müller-Lyer illusion a practice effect—as though subjects learned to avoid an error. What actually happens is, of course, that one distortion is gradually cancelled by another.[32]

Our explanation is not weakened by the fact that during measurements one of the distances in the illusion pattern is objectively varied, shortened or lengthened. Such variations do not essentially alter the condition to which the explanation refers. The angles on one side still "surround" an area which is, therefore, more strongly satiated than the other area. Incidentally, it is not necessary to assume that during the measurements the subject steadily fixates a given point. Even if he moves his eyes back and forth about the center of the figure, satiation of the two areas will remain statistically unequal.

So far we have presupposed that it is a given pattern which is shown, with minor variations, in a set of experiments. But the danger of an unrecognized influence of satiation is not restricted to this condition. Suppose that an experimenter decides to investigate how a certain visual fact is affected by *major* variations of a pattern. In order to eliminate the effects of both practice and fatigue he will give his various patterns in a nicely balanced sequence, and he will determine how the fact in question behaves in each case. Naturally, however, so far as his experimental variations allow he will keep the retinal position of the patterns constant. How does satiation operate under these circumstances? The first pattern which is being used will establish its particular satiation. To this will be added satiation by the second pattern, and so forth, until in continued experimentation and cyclic presentation of the patterns the satiation of the field becomes a mixture of the satiating effects of *all* patterns. If the investigated fact belongs to those which are strongly affected by satiation, the dependence of this fact upon certain variations of the pattern—the actual issue—may thus be obliterated. For under the influence of such a compound satiation individual results will no longer correspond to the particular conditions to which they are referred by the experimenter. Rather, all results will be strongly determined by a more or less constant factor to which all patterns contribute. As a consequence a visual fact may easily appear to be independent of certain variations of conditions although if overlapping satiation effects were absent an influence would be discovered. Roughly speaking, compound satiation will, therefore, make for negative results. We cannot help feeling that this argument may apply to recent experiments of Craik and Zangwill (*cf.* p. 274). These authors did not find that certain variations of configuration affected a threshold in the way which could be expected on theoretical grounds. But their experimental procedure established precisely the condition which we have just discussed.

Early in our investigation we discovered that in some instances the effects of self-satiation were by no means conspicuous (*cf.* pp. 271 and 315). But if our interpretation of Judd's experiments is correct that factor must be most effective in the Müller-Lyer pattern. This raises the question under what circumstances self-satiation has little influence on figures, and under what others it causes great distortions.

Let us first consider an argument according to which, if our theory of figural after-effects is accepted, figures can never be distorted by self-satiation. In immediate application to the

[32] Certain further illusions which are also weakened by repeated inspection will be discussed in the next section. Here we will mention only that in our observation prolonged inspection has the same effect upon Sander's illusion as it has on the Müller-Lyer illusion. When after the inspection period the lines which cause the illusion are omitted from Sander's pattern, in other words, when only the isosceles triangle consisting of the two equal diagonals and the connecting upper horizontal is shown, the diagonals are found to differ in length. The direction of the difference is the opposite of the one observed in the illusion.

Müller-Lyer pattern, we said in Chapter Six (p. 337) that a *T*-line cannot recede from an *I*-line with which it coincides. Now, if we maintain that in the Müller-Lyer figure angles are displaced by self-satiation of the figure, we clearly refer to a case in which "*T*-objects coincide with parts of the *I*-figure." How, then, can the angles be displaced? How, for that matter, can parts of any inspected figure be displaced by their own satiating action?

The answer is that in our former argument we dealt with coinciding *I*- and *T*-lines which were *straight* lines on a homogeneous ground. Under these conditions the satiated area extended symmetrically on both sides of the *T*- (or *I*-) line. It was the *symmetry* of this situation rather than the coincidence as such which prevented any displacement of the line. Therefore, if a given pattern causes *asymmetrical* satiation with regard to the location of one of its own parts, our theory demands that such a part be displaced toward the less satiated side.

This is the rule which all displacements and distortions by self-satiation follow. From the same rule we can also predict in what patterns the changes will be small, and in what others they must be conspicuous. We have just seen that self-satiation cannot displace a straight line in a strictly homogeneous environment. In this situation symmetry of satiation also prevents any *distortion* of the line. A circle is a dubious case. For although a circle satiates its interior more strongly than its environment, the satiated *area* is greater on the outside. Here no safe prediction is possible in the absence of mathematical analysis; and in actual observation some subjects find that after prolonged satiation the circle is a trifle smaller while for others it does not change its size (*cf.* p. 271). Gibson's curved line admits of some distortion by self-satiation because its two sides are not quite symmetrically affected; and so does his bent line. But for more striking effects more strongly asymmetrical situations are needed. As an example we recommend a solid or an outline square having vertical and horizontal straight boundaries. If the center of such a square is steadily fixated the edges of the figure do not remain straight lines. They are gradually transformed into curves which are concave toward the outside. In other words, the corners of the square begin to protrude from its bulk. This change follows from the fact that about the corners of the square self-satiation is as asymmetrical as it is about the angles of the Müller-Lyer pattern. Within the corners of a square the figure current is highly concentrated and, therefore, self-satiation is particularly intense. Opposite the corners, on the outside, just the contrary is the case because here the lines of flow diverge. Thus, each corner lies between a region of intense satiation and another region which is much less affected. As a result the corners soon appear displaced in the latter direction, i.e., toward the outside, because self-satiation shifts the current at the corners in this direction. Similarly, if two large and equal circles are separated by a short distance they will after some inspection assume a flattened appearance where they face each other. For, the narrow space between them is more strongly satiated than their interior so that the adjacent parts of their outlines must recede from that area. Incidentally, when once such a pair of circles was used as an *I*-figure we were surprised to see how rapidly the deformation developed. In fact, we were not sure whether it was ever entirely absent. In the next section we shall come back to this point.

It does not seem to us advisable to give the distortions of figures by self-satiation the name adaptation. It is true that when a curve serves as the inspection object as in Gibson's observations, it may plausibly be said to undergo adaptation in becoming less curved. Here the change is in the direction of a norm. But few will be inclined to use that term when the straight edges of a square become curves, and when its corners begin to point more strongly outward. Such a distortion of the square will rather appear as the opposite of an adaptation. As we see the situation, the difficulty arises from the fact that changes by self-satiation cannot adequately be considered in terms of the distorted figures alone. There is, of course, a strong adaptive trend in *satiation*. This factor seems always to operate in the direction of equalization and uniformity. However, its tendency to prevent inhomogeneities and extreme conditions refers more to the state of the whole medium in which certain patterns are given than to the appearance of these patterns themselves. It is quite possible that self-satiation does not distort certain figures, that it simplifies the shape of others, but that it emphasizes given peculiarities in a third group. Irrespective of such figural differences the direction of the changes will always be the one by

which the condition of the affected tissue as a whole is rendered more even.

Apart from these considerations it would imply too narrow a view of the situation if the effects of self-satiation were defined with reference to the distortion of figures. A single straight line, we repeat, is never distorted by its satiating action. And yet the principle that makes for various distortions in other cases is also operative in this instance: as satiation develops it soon limits itself because it weakens the satiating current. At the same time, as the current now spreads into less affected regions, satiation becomes more uniform in space. This is self-satiation in the absence of figural distortions, and it is also adaptation. Therefore, if the term adaptation is to be used it is plainly such changes of the medium rather than figural changes to which the name should be given.

III

Our report would not be complete if we were to omit the discussion of a last topic. Nobody who is acquainted with problems of vision can have read the preceding chapters without being reminded of certain facts which we have so far not considered. Some familiar illusions[33] are so similar to distortions which have here been described that we must ask ourselves whether this similarity is only accidental.

Figural after-effects consist mainly of displacements. *T*-objects either simply *recede* from satiated regions or, if they are in an oblique position with regard to those regions, they *turn* as they recede. Let us first consider simple displacements including the size effects which result when, in receding from satiated areas, the parts of a *T*-object are driven away from or toward each other.

If a straight line has been inspected two parallel *T*-lines which are then shown in a symmetrical position with regard to the *I*-line will appear too far apart, since both recede from this line. But if we show the three lines simultaneously the distance between the two outside lines tends to be slightly too great at once, i. e., apparently in the absence of satiation (26).[34] In case the distances in this pattern are fairly large we can increase the effect by interpolating a few more parallel lines between the ones first given. We thus obtain what is commonly called the illusion of filled and unfilled space because the expansion of the set of parallels will naturally be demonstrated by comparison with only two lines which delimit an empty area and are objectively as far apart as the lines at the two ends of the set.

When we investigated the displacements which *T*-objects suffer in the neighborhood of *I*-objects we discovered the distance paradox. Within limits the displacements—and corresponding size effects—were greater as the distance between the *I*- and the *T*-objects grew. We showed, for instance, within a previously inspected angle two equal and parallel *T*-lines which were vertical to the bisector of the angle (*cf.* Fig. 28, p. 288). Then the line nearer the open side of the angle, i.e., the line with ends farther distant from the sides of the angle, shrank more than its partner. But if the same angle is shown together with the two lines we obtain Ponzo's illusion in which the length of the lines exhibits a similar difference immediately.

Now, when such distortions occur in the form of figural after-effects they are *greatest* when the distance between the *I*-lines and the *T*-lines has a certain size. The same holds for the illusions, in which *I*-lines and *T*-lines are combined. Thus in our first example a varying number of parallel lines may be interpolated between two given lines. The authors who investigated the illusion of filled and unfilled space report that with a constant objective size of the total distance the expansion is maximal for a certain number of interpolated lines, in other words, for a certain distance between successive lines (14, 18, 26). With Ponzo's illusion a distance of maximal displacement appears in measurements which Sickles has recently reported (23): when one of the two parallel lines in this pattern is kept as a standard in a constant position, namely, deep within the angle, the shortening of the other line depends upon its location. This line is shortest when its ends lie at a certain distance from the angle. It is less shortened when the distance is either greater or smaller.[35]

[33] As to the sense in which we use this term see p. 349, footnote.

[34] Concerning certain exceptions *cf.* 14, 18, 26.

[35] The author proposes an electromagnetic theory of Ponzo's illusion. According to him the sides of the angle are associated with currents—which seems to agree with our theory of figure currents. But then he assumes that the length of a line within the angle decreases as the magnetic field of those currents at the place of the line weakens. It is not explained why the size of lines should thus depend upon a magnetic field. Moreover, the theory is contradicted by the authors's own data which clearly show a

The parallelism between displacements which are figural after-effects and displacements in illusions is not restricted to these particular instances. A considerable number of our Figures—which are combinations of *I*- and *T*-patterns—show immediate distortions of the same kind as are observed in corresponding figural after-effects (*cf.* p. 286, footnote).

We are assuming that the *turning* of *T*-lines in the neighborhood of *I*-lines follows from the same principles as govern simple displacements. Therefore, we shall be inclined to predict that figural after-effects of this kind will also have counterparts which are illusions in the usual sense. Indeed, in Ponzo's illusion we may connect the ends of the parallels by straight horizontal lines so that a square or an oblong is formed. We know that, if such a rectangle is shown after inspection of the angle, its horizontal edges converge toward the open side of the angle. But they do so also if the angle and the rectangle are shown together, although this immediate illusion is less conspicuous than the figural after-effect (*cf.* p. 285).

At this point a sweeping generalization may at first appear quite natural. We remember that the figural after-effect to which we have just referred demonstrates a general principle: Small angles between *I*-lines and *T*-lines always grow. But it has long ago been formulated as a rule that in a number of well-known illusions small angles also grow. Not only the Poggendorff illusion may seem to be interpretable in this fashion but also the distortions which the Zöllner- and the Hering-pattern exhibit. Again, if sets of straight lines pass through an outline circle this figure may be deformed in various ways. The deformations are analogous to the one observed in the Hering pattern. Lastly, Gibson observed several immediate illusions when he combined the *I*-patterns and *T*-lines of his figural after-effects in one drawing. These illusions, which he called contrast effects, had the direction of the after-effects. If our interpretation of his after-effects is correct they also follow the rule that small angles between *I*-lines and *T*-lines grow. But since the corresponding illusions have the same direction they, too, can be subsumed under this rule.

It is one thing to grant that such a formal analogy exists; but as Gibson rightly remarked, it is quite another thing to maintain that the figural after-effects and the immediate illusions are functionally of the same nature. We have to consider the possibility that in some cases the resemblance between the two phenomena is superficial and accidental.

There are at least three factors which may cause illusions. In the first place, visual patterns may be distorted for histological reasons. In section I of this chapter we did not maintain that permanent satiation transforms the visual cortex into a *completely* homogeneous medium. Secondly, distortions may follow from the fact that figure currents are relationally maintained. If to a given pattern further lines are added the flow which pervades that pattern will generally assume a somewhat changed distribution. When functional interrelations within the pattern are thus altered, noticeable distortions may follow. Such distortions would, of course, be independent of any satiation effects; in other words, they would be caused solely by the non-additive way in which figure currents, like all other currents, are distributed. The third possibility is the one which is suggested by the similarity between certain illusions and figural after-effects, i.e., distortions caused by satiation. The only way in which this possibility could be realized is as follows: In certain patterns which may be regarded as combinations of *I*- and *T*-lines self-satiation would develop so rapidly that distortions which are analogous to figural after-effects appear practically at once.

For the most part, illusion patterns are not constructed on clear theoretical grounds. Therefore, a given pattern need not demonstrate the operation of a single principle. Rather, in a given instance various factors may contribute to an effect which is unitary only in appearance. Obviously, this is a further cause for caution in any interpretation of the illusions. None the less, there may sometimes be good reasons for mainly attributing a certain distortion to one of the factors which have just been enumerated. The Vertical-Horizontal illusion, for instance, is clearly not caused by conditions which reside in any particular configurations *per se*. Whatever the ultimate explanation of this distortion may be, some histological factor must be involved. No explanation of this type is possible in the case of the Müller-Lyer illusion. Nor do we see any way of interpreting this distortion in terms of satiation. Rather, the very structure of the Müller-Lyer pattern suggests the hypothesis that functional interrelations across the two crucial distances in this figure are directly

maximum of the illusion at a certain location of the variable line. No such maximum can be derived from his theory.

affected by the presence of the arrow-heads. We are inclined to believe that in a similar fashion direct relational determination is mainly responsible for Sander's illusion.

Now, among the well-known illusions is there one to which an interpretation in terms of self-satiation would be likely to apply? A particularly simple pattern is that of the Poggendorff illusion (Fig. 72). It certainly seems at first

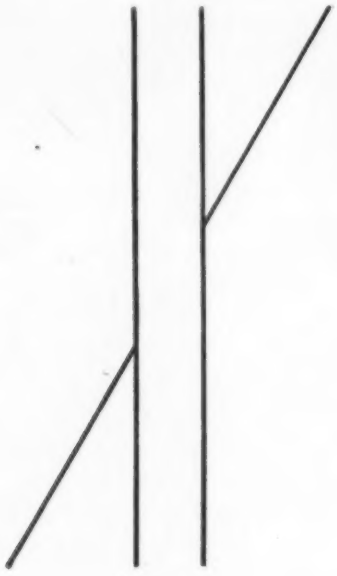


FIG. 72

an attractive thesis that in this figure the two parts of the interrupted oblique line are not seen as continuations of each other because, as a result of rapid self-satiation, each turns away from the vertical with which it is in contact. If this were the case "small angles would grow" in this pattern just as they do in figural after-effects. But we cannot be sure that this is the right interpretation of the illusion because the observed distortion allows also of a different interpretation: if, without turning, both parts of the interrupted line move toward the side of their acute angle, one upward and the other downward, then the same effect will be obtained. In order to decide which alternative is actually realized we have merely to add a neutral landmark, for instance, a *continuous* line which, at some distance, lies parallel to the interrupted line. It will be seen that in this pattern there is only the faintest, if any, indication that the parts of the interrupted line are turned. On the other hand, it is perfectly clear that these parts have not the same distance from the control line; and the direction of the difference is the one which corresponds to the illusion. Thus we learn that in an illusion the rule about the growth of small angles may *seem* to apply while actually there is no connection between that rule and the nature of the illusion. Any serious consideration of such instances as the Zöllner- and the Hering-illusion will lead to similar doubts with regard to these patterns. Even if in some way self-satiation may contribute to such distortions—we have no right to negate this possibility entirely—it seems doubtful whether an increase of small angles by that factor is really the decisive point. On the whole, we believe, the time has not yet come for an explanation of these and related illusions.

This leaves us with a far more limited group of facts which we enumerated in the beginning of this section: the illusion of filled and unfilled space, Ponzo's illusion, and modifications of this distortion in which lines actually turn so that small angles grow. It seems that in these instances the distortions are just those which self-satiation would establish.

But apart from the direction which displacements exhibit in these illusions, their size and their temporal characteristics will have to be considered. It would lead to a lengthy argument if we were now to discuss the less important question of size. It is, however, plainly the crucial problem of our theoretical situation whether self-satiation can develop so fast that its effects appear to a subject as immediately given. For, surely, these illusions do not gradually arise under prolonged inspection.[36]

In this connection it would be convenient to to know precisely how early the presence of satiation can be demonstrated when the appearance of a figural after-effect is taken as a criterion. We have had no difficulty in verifying Gibson's statement that with some patterns an inspection period of five seconds is quite sufficient. But since it seems unlikely that no satiation occurs during shorter periods we made a further test in which the indirect technique of summation was used. On a screen a trapezoid like that of Figure 55 (p. 306), but solid dark on a brighter ground, was repeatedly projected for periods of half a second each. Between exposures the screen was dark for a second and a half. No after-images developed, for in our opaque projec-

[36] Actually, the illusion of filled and unfilled space *decreases* under conditions of prolonged inspection.

tion the brightness of figure and ground differed but little. When after ten exposures (five seconds altogether) two parallel lines like those of Figure 55 were shown as a *T*-pattern, the usual divergence or convergence, opposite to that of the trapezoid, was clearly established. The experiment proves that within one-half of a second a figure process causes a certain degree of satiation even if that process is quite weak. There could be no summation by repetition unless even the first short exposure contributed its share to the final effect. It confirms our other experiences in this field that with an exposure time of only one-half of a second summation occurs across empty intervals of three times that length.

The experiment was repeated when the exposure time was reduced to one-fourth of a second while now the dark intervals amounted to seven times that period. Under these conditions a total exposure time of even less than five seconds (twenty exposures) sufficed to give a clear result. It seems quite probable that if individual exposure times were further decreased we should still obtain after-effects by summation. But for the present it suffices to know that figure processes cause satiation effects within a fraction of a second.

Now, the fact that satiation as such develops so early does not prove that under the conditions of self-satiation the same or even shorter periods can cause overt distortions. We have not found them capable of doing so when the appearance of figural after-effects is used as a criterion. But then, the circumstances under which self-satiation distorts a pattern are not directly comparable to those under which we test for corresponding figural after-effects. To be sure, if the illusions now under consideration are caused by self-satiation, it is satiation of the areas between certain lines which gives rise to the distortions. To this extent conditions are the same in both cases. On the other hand, there is a good reason for assuming that self-satiation affects such an area more quickly than this region is affected when we demonstrate the corresponding after-effects. For in the latter case, in which an *I*- and a *T*-object are shown in succession, lines only on one side of the crucial area contribute to the satiation which develops during the inspection period. If both objects are simultaneously given, as they are in an illusion pattern, the same area is subject to satiation from both sides at the same time. Evidently, in this situation a certain degree of satiation will be reached considerably faster than it is in the other case. Nor is this the only factor that must make for a more rapid appearance of overt effects when the whole pattern is shown simultaneously. While physiological evidence proves that the electrotonic condition as such is extremely persistent, it also tends to show that the highest level of electrotonus which a polarizing current has just established is almost immediately lost when the current is interrupted. Thus, if satiation is identical with electrotonus our experiments on figural after-effects cannot be regarded as tests under optimal conditions. It takes an appreciable time to substitute a *T*-object for an *I*-object. Even if after inspection the subject merely looks from an *I*-screen to a *T*-screen this act involves a certain loss of time. And while such losses may be irrelevant when after-effects are observed after long inspection periods, they may noticeably affect results when we are concerned with the *first* appearance of *any* effect. Since this difficulty is absent in the case of self-satiation any distortion can here show under optimal conditions as to time. It is, therefore, not impossible that certain effects of self-satiation appear practically at once.[37]

All this is far from constituting a proof for the thesis that rapid self-satiation causes the illusions which we have just been discussing. We cannot say more than that the thesis deserves further examination. This examination, however, will have to be thorough. For, if the thesis were to be accepted it would immediately lead to further consequences. An influence of self-satiation upon the immediate appearance of figures could hardly be restricted to a few particular patterns. If, for instance, the size of a set of parallel lines is increased by that factor, why should the size of other patterns not be subject to the same influence? From this point of view, most figures should be enlarged by immediate satiation, for it is always the

[37] Since this was written a further experiment has shown that if *no* time elapses between the disappearance of the *I*-object and the appearance of the *T*-pattern, an after-effect may indeed be clearly visible after an inspection period of only two seconds. With some subjects, it seems, even a single second suffices. It must, however, be remarked that under these circumstances the after-effect is quite short-lived. One can actually see how the *T*-pattern returns to its normal shape. (The experiment was again done with a trapezoid, and two parallel lines as *T*-objects. The trapezoid was projected upon a screen, and when it disappeared the *T*-lines were projected in the proper place by means of a second lantern.)

interior of circumscribed figures in which self-satiation is most strongly and most rapidly established. Obviously, this is a consequence which cannot be accepted unless it be supported by the most convincing evidence. Actually, our principal reason for mentioning this possibility consists in the fact that at present it offers the only way of explaining certain observations which have been described in Chapter One. Consider Figure 18 (p. 280). Here the distance between two small *T*-squares, which lie in a large *I*-square and both equally near its edge, appears smaller than the objectively equal distance between two other such squares farther in the interior of the *I*-square. This difference can hardly be interpreted in terms of displacements which the former *T*-squares suffer in the neighborhood of *I*-contours. Even less would such an interpretation fit the case of Figure 19 (p. 280). Rather it seems that if two objects lie *in a homogeneously satiated region* their distance is shorter than that of similar objects in a neutral region. So far we have offered no explanation for this fact, which has been verified in numerous instances. Now the figure currents of *T*-objects in a homogeneously satiated area must be weakened. For a moment one might, therefore, think that the distance between two objects depends upon the absolute intensity of their figure currents. But this view is incompatible with our assumptions about the correlate of visual distance. For, when the figure currents of two objects are weakened or intensified this change is just as strong on both sides of the objects as it is between them. In other words, the proportion of the flow inside and outside is not altered. What then causes the distance between two objects to shrink when they lie in a homogeneously satiated region? The only answer which we have been able to discover is this: The distance between two such objects does not actually shrink. Rather, this distance *fails to grow* under the influence of immediate self-satiation. In a homogeneously and strongly satiated area not much further satiation can occur; but the distance between two comparison objects in a neutral area *would* be increased by that factor, inasmuch as immediate satiation of the area between these objects would be greater than satiation beyond the objects. Therefore, if immediate self-satiation is a fact, this fact explains the observation to which we have just referred.[38]

[38] Occasionally the same factor may have contributed to distortions the main cause of which is the tendency of *T*-objects to recede from affected areas.

There is still another finding to which an explanation in terms of immediate self-satiation would be applicable. On pp. 330 f. we described an experiment in which the *I*-pattern consisted of two geometrically equal outline oblongs, one black, the other light grey, both on the same white background. We were surprised to see that these oblongs differed in size *on first inspection*: the one with the black outline was considerably larger. Similarly, if of four outline squares on a white ground and in a regular distribution the two on one side were black and the two others light grey, the distance between the black squares was at first sight greater than that between the grey squares (Fig. 73).[39] One is again

FIG. 73

tempted to assume that it is the greater intensity of the figure currents of the black squares which makes their distance appear larger. As a matter of fact, however, neither an increase nor a decrease of their distance can thus be derived. For, as we mentioned before, the currents of the black squares are stronger not only in the area between these squares but also, and in the same proportion, beyond them.

The situation is different if we consider the possibility of extremely rapid self-satiation. Satiation would develop more quickly between two squares than beyond them; obviously, therefore, their distance would be increased. And, since the figure currents of the black

[39] For a convincing demonstration the brightness difference ought to be greater than it is in Figure 73.

squares must be much stronger than those of the pale squares, satiation would enlarge the distance between the former immediately, while this effect would be delayed in the case of the latter. One observation in particular renders this interpretation fairly plausible. The difference between the two distances is not symmetrical. While it is perfectly clear that the lower squares lie at different heights, the corresponding difference between the upper squares is much less conspicuous. This is once more the asymmetry with which we are familiar from many figural after-effects. Undoubtedly, its appearance in the present observation argues for an explanation in terms of immediate satiation. Nevertheless, we insist, final judgment ought to be reserved until further examinations have been performed.

REFERENCES

1. Bales, J. F., and Follansbee, G. L. 1935. The after-effect of the perception of curved lines. *Jour. Exper. Psychol.* **18**: 499–503.
2. Brown, G. G. 1928. Perception of depth with disoriented vision. *Brit. Jour. Psychol.* **19**: 117–146.
3. Craik, K. J. W., and Zangwill, O. L. 1939. Observations relating to the threshold of a small figure within the contour of a closed-line figure. *Brit. Jour. Psychol.* **30**: 139–150.
4. Ebbecke, U. 1933. Zur Lehre vom Elektrotonus. *Ergebnisse Physiol. u. exper. Pharmakol.* **35**: 756–826.
5. Gerard, R. W., and Libet, B. 1940. The control of normal and "convulsive" brain potentials. *Amer. Jour. Psychiat.* **96**: 1125–1153.
6. Gerard, R. W. 1941. The interaction of neurones. *Ohio Jour. Sci.*, **41**: 160–172.
7. Gibson, J. J. 1933. Adaptation, after-effect, and contrast in the perception of curved lines. *Jour. Exper. Psychol.* **16**: 1–31.
8. ——. 1934. Vertical and horizontal orientation in visual perception (Abstract). *Psychol. Bull.* **31**: 739.
9. —— and Radner, M. 1937. Adaptation, after-effect, and contrast in the perception of tilted lines. I. Quantitative studies. *Jour. Exper. Psychol.* **20**: 453–467.
10. Gibson, J. J. 1937. Adaptation, after-effect, and contrast in the perception of tilted lines. II. Simultaneous contrast and the areal restriction of the after-effect. *Jour. Exper. Psychol.* **20**: 553–569.
11. —— and Mowrer, O. H. 1938. Determinants of the perceived vertical and horizontal. *Psychol. Rev.* **45**: 300–323.
12. Henle, M. 1942. An experimental investigation of past experience as a determinant of visual form perception. *Jour. Exper. Psychol.* **30**: 1–21.
13. Kleint, H. 1936. Versuche über die Wahrnehmung. *Zeitschr. Psychol.* **138**: 1–34.
14. Knox, H. W. 1894. On the quantitative determination of an optical illusion. *Amer. Jour. Psychol.* **6**: 413–420.
15. Köhler, W. 1920. Die physischen Gestalten in Ruhe und im stationären Zustand. Braunschweig, Vieweg.
16. ——. 1938. The place of value in a world of facts. Liveright.
17. ——. 1940. Dynamics in psychology. Liveright.
18. Lewis, E. O. 1912. The illusion of filled and unfilled space. *Brit. Jour. Psychol.* **5**: 36–51.
19. Libet, B., and Gerard, R. W. 1941. Steady potential fields and neurone activity. *Jour. Neurophysiol.* **4**: 438–455.
20. Lorente de Nó, R. 1939. Transmission of impulses through cranial motor nuclei. *Jour. Neurophysiol.* **2**: 402–464.
21. Ponzo, M. 1913. Rapports entre quelque illusions visuelles de contraste angulaire et l'appréciation de grandeur des astres à l'horizon. *Arch. Ital. Biol.* **58**: 327–329.
22. Rubin, E. 1921. Visuell wahrgenommene Figuren. Copenhagen, Gyldendalske.
23. Sickles, W. R. 1942. Experimental evidence for the electrical character of visual fields derived from a quantitative analysis of the Ponzo illusion. *Jour. Exper. Psychol.* **30**: 84–91.
24. —— and Hartmann, G. W. 1942. The theory of order. *Psychol. Rev.* **49**: 403–421.
25. Smith, K. U., and Akelaitis, A. J. 1942. Studies on the corpus callosum. I. Laterality in behavior and bilateral motor organization in man before and after section of the corpus callosum. *Arch. Neurol. and Psychiat.* **47**: 519–543.
26. Spiegel, H. G. 1937. Über den Einfluss des Zwischenfeldes auf gesehene Abstände. *Psychol. Forsch.* **21**: 327–383.
27. Verhoeff, F. H. 1925. A theory of binocular perspective. *Amer. Jour. Physiol. Opt.* **6**: 436.
28. Vernon, M. D. 1934. The perception of inclined lines. *Brit. Jour. Psychol.* **25**: 186–196.
29. Wertheimer, M. 1912. Experimentelle Studien über das Sehen von Bewegung. *Zeitschr. Psychol.* **61**: 161–265.
30. Woodworth, R. S. 1938. Experimental psychology. Henry Holt.

"OF THE PEOPLE, BY THE PEOPLE, FOR THE PEOPLE"

JOHN L. HANEY

President (ret.), Central High School of Philadelphia

(*Read April 22, 1944*)

PROBABLY no other event in our national annals has evoked such a variety of contradictory legends as the delivery of Lincoln's Gettysburg Address. This circumstance is the more remarkable because the event did not occur in an early Colonial period from which few authentic sources have survived, but at a relatively late period in national history, with numerous reporters present and an Associated Press dispatch going by telegraph to newspapers throughout the country. The ceremonies were attended by members of the President's Cabinet, by notable military and political leaders, and by hundreds of other prominent persons, yet the conflicting testimony of those who were present on that historic afternoon is, in effect, an astonishing commentary on the reliability of witnesses.

The first group of legends sprang up with regard to the composition of the Gettysburg Address. The varying information on that subject is as follows: (1) it was carefully written in ink at Washington on "Executive Mansion" letter paper; (2) it was hastily jotted down in pencil on the train en route to Gettysburg and was written on (*a*) a sheet of rough foolscap; (*b*) the back of an envelope; (*c*) rough wrapping paper picked up from the floor of the car; (*d*) a piece of pasteboard held on the knee; (3) it was carefully planned and written at Gettysburg in the home of David Wills on the evening of November 18 on (*a*) wide-lined foolscap paper; (*b*) a sort of bluish paper; (*c*) a long yellow envelope; (4) it was hastily composed on a scrap of paper held on the knee during the exercises while Edward Everett was delivering his oration; (5) it was not written at all, but was spoken extemporaneously without notes of any sort.[1]

The second group of legends concerns Lincoln's use of a manuscript while he was delivering the Address. Here we may learn that he held: (1) a card; (2) a slip of paper; (3) several slips of paper; (4) a sheet of paper; (5) several sheets; (6) no manuscript. By way of variety, (7) he held the manuscript close to his face throughout delivery; (8) he referred to the manuscript only occasionally; (9) he drew the manuscript from his pocket, looked at it briefly, and then put it back.[2]

The third group of legends refers to Lincoln's voice. It was: (1) clear, ringing, resonant, vibrating; (2) loud; (3) high tenor; (4) thin and high, but carried well; (5) high clarion tones; (6) deep, powerful; (7) far-reaching; (8) forceful, articulate, could be heard by all; (9) could not be heard, he spoke so low.[3]

The fourth group is associated with the audience's reception of the Address. The testimony of those who were present is to the effect that it was received: (1) with indifference on the part of the tired and inattentive spectators; (2) in silence, with no applause during or after the delivery; (3) with applause in the middle and at the end; (4) with tears, sobs, and cheers; (5) with wild and lengthened excitement; (6) with remarkable enthusiasm; (7) with loud outbursts of applause; (8) with roars of applause; (9) with hurricanes of applause.[4]

The fifth group concerns the popular opinion of the Gettysburg Address immediately after its delivery. The evidence shows: (1) that it was soon proclaimed as a masterpiece of oratory; (2) that it was regarded as a flat failure by Lincoln as well as others; (3) that no Americans appreciated its worth until it was praised by various British authorities several years later; (4) that recognition was slow and did not become general until a much later period.[5]

The sixth group is associated with the phrase "of the people, by the people, for the people," which, in spite of all that has been printed to the contrary, is still widely regarded as Lincoln's

[1] Barton, 66, 204, 67, 164, 153, 173, 64, 171, 191, 68, 69, 205, 167, 172, 163.

[2] Barton, 65, 80, 186, 166, 183, 194, 69, 187, 190, 202, 154, 177, 184.

[3] Barton, 154, 174, 182, 166, 80, 153, 202, 192, 188, 178, 156.

[4] Barton, 89, 163, 91, 154, 173, 188, 192, 201, 176, 165, 194, 181, 67, 90, 170.

[5] Barton, 116, 121, 194, 90, 114, 199, 201, 124; Carr, 80, 81.

own creation and is as definitely identified with him as "all men are created equal" is with Jefferson.

Ordinarily, we would assume that the more we read about a historical event, the more we know. In the case of Lincoln's Gettysburg Address, apparently the more we read the less we know. If we were interested, for example, in so trivial a matter as the weather in Gettysburg on November 19, 1863, we could learn, if we read enough, that: (1) it was a bright, clear day, with the sun shining in glorious splendor; (2) it was a sombre, autumn day, enveloped in gloom; Nature seemed to veil her face in sorrow; (3) in the shadow of the overhanging clouds, the darkness became absolutely funereal; (4) it rained.[6]

If this welter of contradictory folklore belonged to the remote past and had been superseded by a careful presentation of facts, all would be well, but, unfortunately, many of the statements summarized above are still repeated in school texts and by discriminating writers. Among the first to give currency to the legends was Isaac N. Arnold, Congressman from Illinois, who, in his *History of Abraham Lincoln and the Overthrow of Slavery* (1866), stated that on the train en route to Gettysburg, Lincoln was notified that he would be expected to make some remarks and in a short time prepared his address. However, Arnold deserves credit for giving us evidence of early American appreciation of the Address, which, he says, "for appropriateness, comprehension, grasp of thought, beauty, the sublime in sentiment and expression, has scarcely its equal in English or American literature."[7]

During the following years, the curious legends took shape and were repeated with little or no question by one writer after another. Mrs. Mary Raymond Shipman Andrews in her story, *The Perfect Tribute* (1906), put Edward Everett on the President's train going to Gettysburg and made so many other historical mistakes that her tale is often spoken of as a rather imperfect tribute. As recently as 1919, Professor Walter C. Bronson, in his widely used *Short History of American Literature*, represented Lincoln as dedicating the Gettysburg Monument instead of the National Cemetery.[8] Even within the past decade, editors of anthologies and school texts generally have perpetuated the various discredited legends. Just a few months ago, Ambassador Joseph C. Grew, in his entertaining *Ten Years in Japan* (1944), concluded his narrative with an account of his address in Japan on Memorial Day, 1942. He told his hearers that he always repeated the Gettysburg Address on Memorial Day and added that the wearied and harassed President had made rough notes for the address on the back of an old envelope while going down to Gettysburg on the train and that night, while in his hotel, he put the speech into shape.[9]

In view of the current widespread wealth of misinformation concerning the Gettysburg Address, it may be well to review briefly the facts concerning its composition and delivery.

The committee in charge of the dedication of the National Cemetery at Gettysburg had selected October 23 for the ceremonies, but at the request of Edward Everett, the orator of the day, who asked for more time to prepare his speech, the exercises were held on November 19. Formal invitations were sent to President Lincoln, his Cabinet, and to other prominent persons. To the surprise of the committee, Lincoln accepted the invitation. Then arose the delicate question whether he should be asked to speak. Some evidently feared that he might tell a few of his questionable stories, or otherwise disregard the amenities of so dignified an occasion, graced by the presence of the greatest of living American orators, who had been President of Harvard College, Governor of Massachusetts, United States Senator, Minister to England, and Secretary of State. However, on November 2, a carefully worded letter was sent to the President, asking him to make "a few appropriate remarks" after Mr. Everett's address.

In spite of his overwhelming duties, Lincoln found time to give some thought to his prospective "remarks." Everett sent him an advance copy of the oration in the form of proof-sheets of the *Boston Globe*. Lincoln read the text carefully and stated that there was no danger of his getting "on the lines" of Everett's address, as what he intended to say would be short. He wrote at least one page of a draft in Washington before he left for Gettysburg on November 18, but he mentioned that he was having trouble with the closing words. He gave further attention to the draft while on the train and again

[6] Barton, 175, 180, 189, 156.

[7] Arnold, Isaac N., *History of Abraham Lincoln and the Overthrow of Slavery*, 422, Chicago, Clarke, 1866.

[8] Bronson, Walter C., *Short History of American Literature*, 277, N. Y., Heath, 1919.

[9] Grew, Joseph C., *Ten Years in Japan*, 531, Simon & Schuster, 1944.

after his arrival at Gettysburg. As late as eleven o'clock that evening, he discussed the text with Secretary of State Seward, who was one of the three Cabinet members in the Presidential party.

When the exercises got under way at the National Cemetery, somewhat belatedly, at noon on November 19, Senator Everett spoke until a few minutes before two o'clock. After the singing of a hymn, came Lincoln's "few appropriate remarks." Ten sentences—two hundred and sixty-eight words. In a little more than two minutes, Lincoln had added to the American Scriptures one of the most generally revered speeches in our national history.

During the return trip to Washington that evening, Wayne MacVeagh, then Chairman of the Republican State Committee of Pennsylvania, said to the President: "The words you spoke will live in the world's language." Lincoln replied: "Oh, you must not say that; you must not be extravagant about it; you are the only person who has such a misconception of what I said." The next day, Everett wrote to Lincoln: "I should be glad if I could flatter myself that I came as near to the central idea of the occasion in two hours as you did in two minutes." Lincoln, in his unaffected reply, wrote: "I am pleased to know that, in your judgment, the little I did say was not entirely a failure."

Lincoln's Gettysburg Address may be regarded as being in three parts. The first sets forth briefly the reason for founding our nation, the second honors the sacrifice of the brave soldiers to preserve that nation, the third urges increased devotion to the great cause for which those sacrifices were made.

The opening sentence, couched largely in short words of almost Biblical simplicity, followed a traditional custom of beginning with a reference to time: "Four score and seven years ago, our fathers brought forth on this continent a new nation conceived in liberty, and dedicated to the proposition that all men are created equal." Scholars have called attention to a curiously similar passage that began a speech made on December 30, 1850, by Robert Toombs, the Georgia Congressman, who became Senator and later the Secretary of the Confederacy: "Sixty-three years ago our fathers joined together to form a more perfect Union and to establish justice. We have now met to put that Government on trial."[10] Somewhat earlier, Lincoln's Biblical phraseology was anticipated by Alexander Beaufort Meek, Southern writer and legislator, who, in a speech at Mobile on July 4, 1848, said: "three score years and ten and two have passed since the establishment of her nationality by the Declaration of Independence."[11]

We need not assume that Lincoln was indebted to either of these sources for the opening words of the Gettysburg Address. Lincoln's choice of "four score and seven" instead of "eighty-seven" was in keeping with the Biblical phrasing of the entire address. In several of his earlier speeches,[12] notably in the debates with Senator Douglas and in his remarks to the citizens who assembled at the White House on July 7, 1863, Lincoln had quoted and had commented on Jefferson's "all men are created equal" from the Declaration of Independence. The sources of that concept have been studied by many investigators.[13] John Locke had written in his *Treatises on Government*, "men being by nature all free, equal, and independent, no one can be put out of this estate and subjected to the political power of another, without his own consent" and "though I have said 'that all men by nature are equal,' I cannot be supposed to understand all sorts of equality."[14] For Jefferson, the writings of Locke were as a political Bible. The basic idea of political equality can be traced at least as far back as Aristotle, but its development came largely by way of the English writers during the Parliamentary struggles of the seventeenth century and not, as is still so generally believed, from the French philosophers of the eighteenth century.

The second part of the Gettysburg Address has been recognized as a lofty expression of a sentiment that dates from the days of Pericles, if not earlier. It is the grateful appreciation of those who benefit by the supreme sacrifice of others. The analogies between Pericles' Funeral Oration over the Athenian soldiers who had fallen in battle and Lincoln's Gettysburg Address were apparently first noted by George

[10] Hertz, *Abraham Lincoln* I: 103.

[11] Barton, 130.

[12] Peoria Speech, Oct. 16, 1854; Letter to George Robertson, Aug. 15, 1855; Letter to Joshua F. Speed, Aug. 24, 1855; Springfield Speech, June 26, 1857; Douglas Debates, Aug.-Oct. 1858.

[13] Friedenwald, Herbert, *The Declaration of Independence*, Macmillan, 1904; Hazelton, John H., *The Declaration of Independence*, Dodd, Mead, 1906; Becker, Carl L., *The Declaration of Independence*, Harcourt, Brace, 1922; Knopf, 1942.

[14] Friedenwald, *Declaration of Independence*, 189, 201.

William Curtis in *Harper's Weekly* as early as April 23, 1864, but the resemblance is general, not specific. Few students have stressed the obvious and more immediate obligations of Lincoln to Everett's oration, which he read carefully and repeatedly during the days preceding the dedication. In fact, when Lincoln went to be photographed on Sunday, November 8, he took with him the Everett text to read between sittings. One photograph reveals the envelope containing the proof-sheets lying on the table beside Lincoln. Everett, as a devoted classical scholar, was steeped in the tradition of our obligation to our soldier dead. He began his oration with an account of the Athenian obsequies for those who had fallen in battle and he quoted the most famous words of Pericles near the end. Lincoln did not have to go further than Everett's text to find the inspiration for the middle portion of his address, in which he expressed, with greater cogency and simplicity, what Everett had written as follows:

> We have assembled, friends, fellow citizens, at the invitation of the Executive of the great central state of Pennsylvania, seconded by the Governors of seventeen other loyal States of the Union, to pay the last tribute of respect to the brave men, who, in the hard fought battles of the first, second and third days of July last, laid down their lives for the country on these hill sides and the plains before us, and whose remains have been gathered into the Cemetery which we consecrate this day. As my eye ranges over the fields whose sods were so lately moistened by the blood of gallant and loyal men, I feel, as never before, how truly it was said of old, that it is sweet and becoming to die for one's country. I feel as never before, how justly, from the dawn of history to the present time, men have paid the homage of their gratitude and admiration to the memory of those who nobly sacrificed their lives, that their fellow-men may live in safety and in honor. And if this tribute were ever due, when, to whom, could it be more justly paid than to those whose last resting place we this day commend to the blessing of Heaven and men?

From this stately and carefully wrought passage of Everett's, couched in the oratorical tradition of the period, Lincoln drew the inspiration for his own matchless prose:

> We are met on a great battle-field of that war. We have come to dedicate a portion of that field as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this. But in a larger sense we cannot dedicate, we cannot consecrate, we cannot hallow, this ground. The brave men, living and dead, who struggled here have consecrated it, far beyond our poor power to add or detract.

Everett, in paying a tribute to the soldier dead, spoke in the spirit of Pericles; Lincoln, by shifting the consecration to the soldier dead, won immortality for his words.

The third part of the Gettysburg Address reveals a degree of climactic eloquence never before achieved by Lincoln and matched only once thereafter in the Second Inaugural Address. It reveals the very essence of true eloquence that sets this Address apart from most of the fustian oratory of our national history. Starting with his memorable words of self-effacement, so rare in high places, he called upon his hearers to dedicate themselves to the obligation confronting them:

> The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us, the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us,—that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion; that we here highly resolve that these dead shall not have died in vain; that this nation, under God, shall have a new birth of freedom; and that government of the people, by the people, for the people shall not perish from the earth.

Interest centers, of course, in the tripart phrase at the end. It did not attract much attention at first, as it was probably accepted as original with Lincoln and as representing a rhetorical flourish to conclude his brief remarks. Later, approximate analogues came to light, but usually in a text that Lincoln had never seen. Several alleged equivalents of Lincoln's words were brought forth, but could not be substantiated. Now it can be stated, however, that one writer anticipated Lincoln's exact phrasing.

Let us examine these analogues briefly in their chronological order.

During the age of Pericles, the tanner Cleon, a blatant, unscrupulous demagogue, is alleged to have said to the men of Athens: "I am in favor of the democracy that shall be democratic, that shall give us the rule, which shall be of the people, by the people, for the people." The man who first discovered this quotation from

Cleon has not been identified. Markens[15] referred to him in 1913 as "a diligent researcher." Other diligent students of later days have sought the quotation in vain. As a matter of fact, Thucydides quotes Cleon as voicing an exactly opposite thought on the incompetency of democracy. While we would not expect so typical a demagogue as Cleon to be consistent, and while we know him mainly through the rough satirizations of Aristophanes, there is no reason for assuming that Cleon was a sort of Abraham Lincoln of ancient Athens.

The second apparently spurious analogue is a quotation alleged to be from the Preface of Wyclif's translation of the Bible (about 1382): "This Bible is for the government of the people, by the people, for the people." Once more the name of the original "diligent researcher" is lacking. The quotation can be traced as far back as 1895, when Ward Hill Lamon, in his *Recollections of Abraham Lincoln*, stated that it could be found "in the preface of the old Wickliffe Bible published A.D. 1324." An edition of the Bible by the very youthful Wyclif well over a century before the invention of printing might well arouse suspicion. Barton stated that "many authorities have declared that Lincoln got the phrase from the Introduction to Wycliff's Bible," and, curiously enough, quotes Hall Caine as stating (about 1925) that a few years previous he had himself found the passage in Wyclif's preface.[16] Barton had the librarians of Washington, New York, and Boston look through their Wyclif Bibles, but they found nothing, so he concluded that Hall Caine was the only man who had been able to find it. However, the citation from Lamon (1895) would prove that at least one investigator had anticipated Sir Thomas Henry Hall Caine in finding a quotation that was not there.[17]

With more confidence we can review some of the earlier forms of the phrase in the later eighteenth century. When Patrick Henry, who had pleaded so earnestly for liberty in the early days of the Revolution, refused to attend the Constitutional Convention in 1787, he opposed the adoption of the Constitution because he foresaw what he considered a "consolidated government" endangering the sovereignty of the States, and he urged instead the establishment of "a government of the people, by the people," thus anticipating two-thirds of Lincoln's phrase.

Dr. Thomas Cooper, English lawyer and philosopher, in his *Some Information Respecting America* (1793), wrote: "the government is of the people and for the people."[18] When Cooper returned to London, he became known to Samuel Taylor Coleridge and Robert Southey, who were then zealously promoting their project to establish a social utopia or Pantisocracy in America. The young dreamers saw much of Dr. Cooper during September, 1794. It was he who recommended "the Susquehannah" as the location for their experiment. Neither enthusiast knew much about America, but both liked the poetic name of the river mentioned by Cooper. Coleridge was a careful reader of Cooper's book and a year later, in his address, *The Plot Discovered* (1795), he wrote: "Governments have assumed many different forms, but in their essence and properties, all possible modes of government are reducible to these three: government *by* the people; government *over* the people, and government *with* the people." He then proceeded to explain these three forms in detail.[19]

On July 11, 1798, President John Adams, in addressing the people of Westmoreland County, Virginia, said: "The declaration that our people are hostile to a government made by themselves, for themselves, and conducted by themselves, is an insult." In that brief characterization of the nature of our government, President Adams anticipated, somewhat lamely, to be sure, the essence of President Lincoln's more famous words.

After the lapse of some twenty years, we reach the first of the really significant prototypes for Lincoln's phrase and possibly the first that was known to Lincoln. Chief Justice John Marshall, in his decision (1819) in the case of McCulloch vs. Maryland, wrote: "The government of the Union is emphatically and truly a government *of* the people. In form and in substance it emanates *from* them. Its powers are granted *by* them, and are to be exercised directly *on* them and *for* their benefit."[20] In that statement, Justice Marshall achieved the distinction of using five prepositions in his generalization; others were content with two or three. Thus, a year later (November 29, 1820), President

[15] Markens, 14.
[16] Barton, 132, 152–155.
[17] Lamon, Ward Hill, *Recollections of Abraham Lincoln*, 176, McClurg, 1895. See also Stevenson, Burton, *Home Book of Quotations*, 431, Dodd, Mead, 1937.
[18] Cooper, 52, 53.
[19] Coleridge, 33, 43.
[20] Wheaton, *Reports of the Cases of the Supreme Court of the United States*, IV: 403, N. Y., 1827.

Monroe wrote in his *Fourth Message to Congress:* "a government which is founded by, administered for, and is supported by the people." Similarly, President Jackson, in his first *Inaugural Address* (1829), said: "As long as our government is administered for the good of the people, and is regulated by their will; as long as it secures to us the rights of persons and of property, liberty of conscience and of the press, it will be worth defending."

A prototype that was certainly familiar to Lincoln was in Daniel Webster's Second Speech on the Foot Resolution—the memorable Reply to Hayne—in which he spoke (January 26, 1830) of "the people's government, made for the people, made by the people, and answerable to the people." This famous speech was used by Lincoln when he was composing his First Inaugural Address. On the other hand, it is most unlikely that Lincoln was familiar with a speech delivered in Olten, Switzerland, in May, 1830, by an orator named Schinz, who said: "All the governments of Switzerland must acknowledge that they are simply from all the people, by all the people, and for all the people." This quotation is of interest because of Schinz's use of "all," which was appropriate only because of the approximately pure democracy prevalent in certain parts of Switzerland. This speech is usually cited by the "diligent researchers," but thus far they have overlooked the fact that it first became known in P. Feddersen's *Geschichte der Schweizerischen Regeneration von 1830 bis 1848*, which was published in Zurich, 1867, some two years after Lincoln's death. Another analogue that was probably unfamiliar to Lincoln was furnished by a Scotchman named James Douglas in *The Advancement of Society* (1830): "the depressed vassal of the old Continent becomes co-legislator and co-ruler, in a government where all power is from the people, and in the people, and for the people."[21]

The next two decades produced only a few anticipations of Lincoln's phrase. When the Reform Bill was introduced in the British Parliament (1831), Lord John Russell, speaking of the House of Commons, called it "a body . . . representing the people, springing from the people, and sympathizing with the people." The guarded character of Lord Russell's words, in contrast with Daniel Webster's forthright statement, emphasizes the slower progress of democracy in Great Britain. A somewhat similar phrase occurs in Joseph Story's *On the Constitution* (1833): "In a government like ours, founded by the people, managed by the people."

In Europe, during the period of political ferment preceding the revolutions of 1848, there were numerous writings on the relations between the people and the State. The pamphlets and the treatises of that age would yield many analogues, all of which were doubtless unfamiliar to Lincoln. Thus, Friedrich Karl von Savigny, in his eight volume work on the *System of Contemporary Roman Law* (1840–1849), made this categorical statement: "Originally and according to nature, the State had its origin in a people, through that people, and for that people." Alphonse de Lamartine, in his *History of the Girondists* (1847), in referring to the ideas of Robespierre, said: "This end was the representative sovereignty of all citizens, concentrated in an election as extensive as themselves, and acting by the people and for the people in an elective council, which shall be all the government." Elsewhere Lamartine stated:

> We preach *Democracy* in vain while Tory and Conservative can point to the opposite side of the Atlantic and say: "There are nineteen millions of the human race free absolutely, every man heir to the throne, governing themselves—the government of all, by all, for all; but instead of being a consistent republic, it is a widespread conspiracy of free men for the enslavement of a nation of another complexion." [22]

This brings us to an analogue that anticipated Lincoln's exact expression except for the use of the word "all" in each of the three phrases. At an Anti-Slavery Convention held in Boston on May 29, 1850, Theodore Parker, the aggressive reformer and abolitionist, said: "There is what I call the American idea . . . this idea demands . . . a democracy,—that is, a government of all the people, by all the people, for all the people." Parker's views were much ahead of his times and he had a definite purpose in mind in including and perhaps stressing the word "all." Four years later, when he addressed the Anti-Slavery Society in Boston on May 13, 1854, he anticipated still more of the Gettysburg Address: "For there is the democratic idea: that all men are endowed by their creator with certain natural rights . . . that they are equal as men . . . and therefore government is to be of all the people,

[21] Douglas, 70.

[22] Lamartine, English trans., Bohn ed., III: 104, 1850.

by all the people, for all the people." Still more interesting is a citation from a sermon delivered by Parker in the Music Hall of Boston on July 4, 1858. On that occasion he varied his phrasing somewhat by saying: "Democracy is direct self-government, over all the people, for all the people, by all the people." Apparently by that time he had decided that "of" was colorless or ambiguous and substituted "over"; he also placed the "by" phrase at the end, regarding it as a climax. The sermon was published as a pamphlet entitled *On the Effect of Slavery on the American People*. Herndon gave a copy of the pamphlet to Lincoln, who not only read it, but underlined with his pencil the passage just cited.

In addition to the three quotations from Theodore Parker, the 1850's produced several other parallels for Lincoln's phrase, but never quite in Lincoln's own words. Lieutenant Matthew Fontaine Maury, oceanographer and naval officer, in a Report to Congress (August, 1851), wrote: "Unlike Europe, there are no disaffected people in this country for a foe to tamper with. The government is by the people, for the people, and with the people. It is the people."[23] The concluding epigram, unique with Lieutenant Maury, should help to keep his memory green. George Thompson, the English anti-slavery leader and orator, in a speech delivered in 1851, is said to have used the expression "government of all, by all, for all." More significant is a passage from a speech by Judge Joel Parker on May 16, 1853, when he declared:

> What did the people of Massachusetts undertake to do when, in 1780, they attempted to make a Constitution? . . . They undertook to solve the problem of self-government, to establish a government, not distinct from, and adverse to, the people, but a government of the people themselves, to be administered by and through the people.[24]

This phrasal maze of Judge Parker's sounds almost like an attempt to match Chief Justice Marshall's masterpiece of the five prepositions.

We now reach the only authenticated analogue in the exact words used by Lincoln in the Gettysburg Address. Late in 1860, a group of young, unorganized abolitionists of Boston arranged to hold some meetings in the Tremont Temple. Very unwisely, they decided to hold their sessions, morning, afternoon and evening, on the anniversary of the execution of John Brown, who had been hanged on December 2, 1859. As that date fell on a Sunday in 1860, they called their meetings for Monday, December 3. The committee in charge, headed by James Redpath, lecture promoter, journalist and reformer, sent invitations to numerous prominent persons. Many of these were in sympathy with the young abolitionists, but disapproved heartily of John Brown. Among those who declined to attend was Governor William F. Packer, of Pennsylvania, whose sharply worded letter to the committee was printed in the Boston *Evening Transcript* on November 28, 1860. Another who declined was Henry Wilson, who had succeeded Edward Everett as Senator from Massachusetts and who later became Vice-President of the United States in Grant's second administration. The letter which Senator Wilson wrote to James Redpath and his associates on November 27, 1860, contained the exact phrase that Lincoln used at Gettysburg almost three years later. It was printed in the *Evening Transcript* on December 4 and in other Boston daily and weekly papers about the same time.[25] In view of its significance in connection with the saga of the Gettysburg Address, it is reprinted in full:

> Natick, Nov. 27, 1860
>
> GENTLEMEN,—I have received your invitation to meet in a Convention to be held in Boston, on the 3d of December, "a number of young men unconnected at this time with any organization," and to address them in reply to the question—"How can American Slavery be abolished?" You say in your note of invitation, "that the anniversary of the death of John Brown, who, on the 2d of December, 1859, was killed for attempting to decide this problem in the mode that he believed to be the most efficient, is an occasion peculiarly appropriate for the discussion of our duty to the race for whom he suffered, and more especially for the unfolding of practical methods for achieving the holy object he desired to attain by his descent on Harper's Ferry." Abhoring slavery in every form,—loving equal and impartial Liberty for all men,—I am ever ready to exercise all the powers of the Constitution of my country to relieve the nation from all connection with, and all responsibility for slavery, by prohibiting it, wherever it exists, or can exist, under the exclusive authority of the Federal Government; and I am also ever ready to use all means sanctioned by law, humanity and religion, to persuade our countrymen of the slaveholding States, to "undo the heavy

[23] Markens, 14.

[24] Hertz, *Abraham Lincoln* I: 152.

[25] Boston *Post*, *Courier*, etc. See also the (weekly) *Liberator*, Nov. 30, Dec. 7 and 14, 1860. Senator Wilson's letter was reprinted in the Dec. 14 issue.

burden," and "let the oppressed go free"; but I am utterly opposed to all appeals, by whomsoever made, to force and violence. Ours is a government of constitutions and laws,—a government of the people, by the people, for the people. Not therefore to the rifle or the pike should the friends of the slave appeal, but to the heart, the conscience, the reason and the enduring interests of the people of the Slave States, upon whom rests the responsibility of slavery in the States.

While I accord to John Brown sincerity of motive, unselfish devotion, and unsurpassed courage, I deplore and condemn his armed invasion of Virginia. Were it in my power to do so, I could not consent to attend your meeting, for I cannot, by an act of mine, thus associate in any degrees the issues concerning slavery with John Brown's lawless descent upon Harper's Ferry.

Yours truly,
HENRY WILSON

To

JAMES REDPATH, ESQ., and others of the Committee

Lincoln had been elected President on November 6, nearly a month before Wilson's letter was printed, but he did not leave Springfield until February 11, 1861. It is unlikely that he became familiar with the text of the letter that anticipated the exact phrasing of the Gettysburg Address.

Lincoln's regard for the Declaration of Independence, from which he drew his inspiration for the first part of the Gettysburg Address, is shown in the speech which he delivered at a flag-raising at Independence Hall, Philadelphia, at six o'clock in the morning on Washington's Birthday, 1861, while he was on his way from Springfield to Washington for his inauguration. He had reached Philadelphia on the afternoon of February 21 and that evening, in his apartment at the Continental Hotel, he had a visit from Allan Pinkerton, the well-known detective, who informed the President-elect of a plot to assassinate him as he passed through Baltimore on the twenty-third. This circumstance gives dramatic significance to certain remarks near the close of Lincoln's Independence Hall Address, in the course of which he said:

All the political sentiments I entertain have been drawn, as far as I have been able to draw them, from the sentiments which originated in and were given to the world from this hall. I have never had a feeling, politically, that did not spring from the sentiments embodied in the Declaration of Independence. . . . It was that which gave promise that in due time the weights would be lifted from the shoulders of all men, and that all should have an equal chance. . . . If this country cannot be saved without giving up that principle, I was about to say I would rather be assassinated on this spot than surrender it. . . . I have said nothing but what I am willing to live by and, if it be the pleasure of Almighty God, to die by.

Within two weeks, at his Inaugural Address on March 4, 1861, Lincoln made a last minute appeal to the South. In it he referred to his "rightful masters, the American people," and stated: "this country, with its institutions, belongs to the people who inhabit it." Again: "Why should there not be a patient confidence in the ultimate justice of the people? Is there any better or equal hope in the world?" Just four months later, in a message to Congress on July 4, President Lincoln wrote: "This issue embraces more than the fate of these United States. It presents to the whole family of man the question whether a constitutional republic, or democracy—a government of the people by the same people—can or cannot maintain its territorial integrity against its own domestic foes." This passage makes clear that Lincoln's concise definition of democratic government was then taking shape in his mind.

The last speech of Lincoln's that had a bearing on the text of the Gettysburg Address—and one that has not been sufficiently emphasized in that connection—was delivered to a group of citizens who had assembled at the White House on the evening of July 7, 1863, to serenade the President in honor of the victories at Gettysburg and Vicksburg a few days before. Lincoln's informal remarks on that occasion reveal several anticipations of the Gettysburg Address:

Fellow Citizens: I am very glad to see you tonight, and yet I will not say I thank you for this call; but I do most sincerely thank Almighty God for the occasion on which you have called. How long ago was it?—eighty-odd years since, on the Fourth of July, for the first time in the history of the world, a nation, by its representatives, assembled and declared, as a self-evident truth, "that all men are created equal" . . . and now on this Fourth of July just passed, when we have a gigantic rebellion, at the bottom of which is an effort to overthrow the principle that all men are created equal, we have the surrender of a most powerful position and army on that very day. I would like to speak in terms of praise due to the many brave officers and soldiers who have fought in the cause of the Union and

liberties of their country from the beginning of the war.

In these words, Lincoln extemporized a sort of prosaic first draft of the superb poetry and exalted sentiment which, a few months later, characterized his "few appropriate remarks" at Gettysburg.

It would be futile to follow the genesis of the Gettysburg Address any further. That message has a place in the hearts of Americans second only to the Declaration of Independence. When William Tyler Page, Clerk of the House of Representatives, on April 3, 1918, formulated the American Creed, which was accepted by the House on behalf of the American people, he began with these words: "I believe in the United States of America, as a government of the people, by the people, for the people, whose just powers are derived from the consent of the governed"—and that is as it should be in the Creed of what Woodrow Wilson once called "the only idealistic nation in the world."

REFERENCES

BARTLETT, JOHN R. 1874. Soldiers' National Cemetery at Gettysburg. Providence Press.

BARTON, WILLIAM E. 1930. Lincoln at Gettysburg. Bobbs, Merrill.

BOOKER, RICHARD. 1941. Abraham Lincoln in periodical literature (1860–1940). Chicago, Fawley-Brost.

BURRAGE, HENRY S. 1906. Gettysburg and Lincoln. Putnam.

CARMICHAEL, ORTON H. 1917. Lincoln's Gettysburg address. Cincinnati, Abingdon Press.

CARR, CLARK E. 1906. Lincoln at Gettysburg. McClurg.

DODGE, DANIEL K. 1924. Abraham Lincoln, master of words. Appleton.

GREEN, SAMUEL A. 1901. President Lincoln's speech at Gettysburg. Boston, Mass. Hist. Soc.

HERTZ, EMANUEL. 1931. Abraham Lincoln, a new portrait. 2 v. Liveright.

JACOBS, HENRY E. 1919. Lincoln's Gettysburg world-message. Phila., United Lutheran Pub. House.

LUHRS, HENRY E. 1938. Lincoln at the Wills house and the Gettysburg address. Shippensburg, Pa., The Lincoln Publishers.

MADIGAN, THOMAS F. 1931. A few appropriate remarks by Abraham Lincoln. N. Y., privately printed.

MARKENS, ISAAC. 1913. Lincoln's masterpiece. N. Y., W. H. Christian, printer.

MOORE, CHARLES. 1927. Lincoln's Gettysburg address and Second Inaugural. Houghton, Mifflin.

ROBINSON, LUTHER EMERSON. 1923. Abraham Lincoln as a man of letters. Putnam.

SANDBURG, CARL. 1939. Abraham Lincoln, the war years. 4 v. Harcourt, Brace.

TAUSEK, JOSEPH. 1933. The true story of the Gettysburg address. N. Y., Dial Press.

DAKOTA TREATMENT OF MURDERERS*

ELLA CARA DELORIA

Columbia University

(*Read April 22, 1944*)

THE Dakota Indians commonly known as Sioux made a clear distinction between the killing of outsiders and the killing of a fellow tribesman. In the one case it was regarded as a legitimate part of warfare while in the other it was murder, and therefore a crime against society and liable to punishment. But what that punishment consisted of has not been generally studied. For that reason it is my intention to give you briefly at this time some interesting material on that question which I have been able to find.

The material is not complete as yet. It comes from but two of the three cultural groups of the Dakotas, Yankton and Teton. I have still to get it from the third group, the Eastern Dakota or Santee. But as far as I have gone I find three distinct methods of dealing with murderers. These I shall discuss in the following order:

1. Immediate reprisal killing of the murderer by a male relative of the slain; peacemaking optional.
2. Trial by ordeal.
3. Adoption of the murderer in place of his victim, by the latter's relatives.

The order is purely arbitrary. I do not aim to imply any consistent progress of the entire tribe, as time went on, towards a more humane treatment of the guilty. On the contrary it comes out quite clear that the method chosen in each instance was dependent upon the mood and character of the injured relatives.

The first method may be easily explained, for it was the simplest. Some angered relative, usually a brother or cousin, took it upon himself to settle the score in behalf of his kinsmen by stabbing or shooting the murderer at once. In such a case, the murderer's relatives generally accepted his act as an inevitable consequence, and did nothing about it. As for the kinsmen of the murdered, they looked upon the man of their number who did the reprisal killing as admirable. An ugly job had been facing them all and he had done it. Naturally they were morally indebted to him because of it. "Our kinsman has soiled his own hands to spare us."

But occasionally it happened that the murderer's relatives were displeased. Perhaps they felt that there had been enough provocation for the murder. That being their attitude, one of them might swear revenge on the man who had done the reprisal killing. Then between these two men there was bad feeling, affecting all the people, for they were all related tribally; if not through blood or marriage, at least through social kinship. But if the two men were of no special importance anyway, nothing was done about their quarrel. "Let them fight it out" seemed to be the council's attitude.

If, on the other hand, either man or both were normally men of peace and good will within the tribe, then something was done. From the augmented council comprising not only the current councilmen but also prominent leaders, two men of good character, who were likely to command the respect of the two men in trouble, were sent to the quarreling ones to persuade them to make peace. The language in which they were commissioned is interesting. The speaker of the council handed each a pipe of peace and said in effect:

"Take this pipe to our unhappy brother. . . . Ask him to accept it. . . . While there is bad feeling in our midst our children cannot be safe, our women cannot be at ease in their hearts. . . . It is every Dakota's responsibility to maintain peace, for peace is our heritage; it is in our name." (N.B. Odakota means peace; the *O* is the locative prefix, and the word literally says: "In a state or condition of Dakota-*ness*." Dakota, the tribal name, means: in friendship, fellowship; allied; in understanding of one another's speech and thought.)

It was both a great honor and a fearful re-

* This paper reports the results of an investigation supported by a grant from the Penrose Fund of The American Philosophical Society.

sponsibility for the ones sent out. Their success was a measure of their influence for good and their worth to the people. Sometimes such a messenger returned unsuccessful, finding a very angry man who refused to listen to his pleading; and then another man was sent. In the end the two at odds were brought to the council tipi, having finally yielded; and there they were caused to smoke the peace pipe in communion, and kindly words were said to them, and they were feasted. Then those tribesmen who were especially desirous of peace and were able to contribute goods, assembled two piles of presents which were given to the two men "to cool off their hearts with them." And it was not the intrinsic value of the gifts that mattered; it was what they symbolized: that, to their friends, the happiness of those two men was more important than mere goods.

The two who had acted as mediators expected and received nothing for their offices. Their reward lay in the added prestige which their success had brought them.

It has been suggested to me, facetiously I think, that it must have been profitable among the Dakotas to become involved in a bitter quarrel and thereby come into great wealth all in a day. But anyone who knows Dakota life knows that no Dakota man would go to so much trouble for mere things. Only sudden anger or grief, and a desire for revenge, or family pride, or some such feeling, could cause a quarrel of this sort. In a society where giving rather than getting was glorified, and material goods were constantly changing hands, where it was considered unsocial to amass worldly goods for oneself, and anybody who did so was immediately suspect, a material motive would hardly answer.

In trial by ordeal which I list second, there appear to have been two different forms for carrying it out. One was called "C'a-ápsil-k'iyapi" (which I shall call the Hurdles) and the other was called "Wayáka Kiyúśkapi" (Captive Released).

Trial by ordeal was resorted to if nobody from the injured family took prompt vengeance by killing the murderer. One of my informants explained it thus: "Occasionally it happened that all the male relatives of the slain were pacifists and would not kill even in retaliation. Such men preferred to let the Great Mystery decide the murderer's fate."[1]

[1] Contrary to the prevailing notion that every Dakota was a killer, meriting the term "cut-throat" given the tribe by their enemies, actually there were men who were opposed to taking human life, who chose life occupations far removed from warfare. Now and then when I ask some old man about a point in war I have had such answers as: "I have always been a man of peace so I do not know. . . Ask my cousin, for he was a fighting man."

Another said: "If it happened that all the male relatives were weak-hearted men (that means cowardly) and were afraid to kill even though their family had been outraged, they turned the matter of punishment over to the council." One informant was in sympathy with such relatives while the other was openly contemptuous of them.

Perhaps we shall get a clearer picture of the ordeals if we have an account of them, so I give now my free translation of the description of White Horse (Teton) as he gave it to me in the native language.

"This my grandfather witnessed, and told to me. A youth had been wantonly slain whose relatives were all gentle, timid people who aimed to keep out of trouble. These said to the council: 'We cannot kill. We are pained, but we cannot kill. Do with the murderer as you will.' And so the council resorted to trial by ordeal, using the Hurdles. This is how it was:

"Four hurdles were set up at reasonable intervals along a course. It was kept clear of people and dogs. It ran in front of the council tipi where the men of importance sat to watch, able to see, because the bottom of the tipi was propped up all around, it being a warm day. Horizontal bars were laid on uprights at a level suited to the man's height, but they were not tied on. It was possible to clear them all without knocking any off—but only with supernatural help. Otherwise, even the movement of air as he passed could knock them off.

"My grandfather said it was a solemn day, oppressive to every one. Especially it was so to the relatives of the murderer. Having no hope of a happy outcome for him, they were already mourning while they stood waiting. The timid relatives of the slain man were also on hand. They were required to stand ready with their bows and arrows and, if the runner failed, that instant they were required to shoot at him together and so kill him; thus all inflicting on him together the punishment that none would undertake alone.

"Men, women, and children thronged about, for this was something everyone dreaded both to see and to miss seeing. The murderer was now

escorted from a nearby tipi to the starting line. He was nude but for a breech cloth; his feet were bare. At the usual signal cry, 'Ho-ka-hé!,' he ran, and cleared three hurdles but failed on the fourth; and then, almost before anyone could see, he was down, with as many arrows sticking into his back as his victim had male relatives. And that was right. The Something-Holy could very well have helped him clear that last hurdle, but he had not. That meant he must die. So nobody could blame those who shot him. It was supernatural justice.

"If ever a murderer survived one of these ordeals, it was because the Something-Holy, the Great Mystery, had exonerated him. He was therefore set free to come and go among the people as formerly. None dared to molest such a one. Whom the Supernatural had helped, no man might presume to judge."

The alternate form of the ordeal, the Captive Released, was substituted whenever the council so ordered. The scene was about the same except that, instead of hurdles being set up, a very spirited horse which till then had run wild and never been ridden was brought in for the murderer to ride, bareback and without reins.[2] Men struggled to hold it, until the man was seated, and then they released it and it went bucking and rearing and whirling about in an effort to throw him. He clung to the mane for dear life, literally, knowing that if and when he fell he would be killed by the arrows of his victim's relatives, perhaps even before he struck the ground. One informant added that the youth of the community took delight in waving blankets and shouting to frighten the horse even more. He said it was generally thought that this form was even harder than the Hurdles.

There is a gruesome variant to these ordeals, or I should better say, an added feature, which I cannot omit. It appears that occasionally it was decided that the passing of the ordeal as already described would not be quite enough; that the murderer ought to pay still more. So they laid the corpse of his victim in the honor-place of the tipi, and arranged the hurdles so that as he cleared the last one he landed within the entrance to the council-tipi. Or if he was undergoing the Captive Released, he was expected to leap from the horse when it neared the place, and enter. The councilmen sat solemn on either side of the room and the murderer was ordered to lie down on the cold naked body and feel it with his entire body; warm limb against cold limb, face against face; and so to learn the enormity of his act: that he had taken what he could never restore. Then a piece of cooked meat was touched to the dead lips and he was caused to eat it; and the pipe, lighted and ready, was inserted between the dead lips and he was caused to smoke it. And so at last he atoned for his deed.

"*They* say this is what happened" is the way my informants spoke of this feature, "*they* say it happened long ago." Nobody could say of it, "Yes, my great-grandfather said his grandfather *saw* it." It almost belongs with the legends.

We have still to consider the third method; that of adopting the murderer in his victim's place.

My best account comes from Antelope, a Yankton Dakota, well known and universally respected for his integrity. A man well over seventy, but keen of memory and forthright of speech, who weighed each word before he spoke it; a most reliable informant. I think once again I shall quote:

"This method of dealing with a murderer was to my mind the finest of all. But it took great self-mastery and generosity, and so it was rarely used, since those are rare qualities.

"Suppose a man had been killed, and his relatives met, to decide what to do. They asked themselves: 'Shall one of us go now and finish off the murderer? Or, shall we turn the matter over to the councils asking for trial by ordeal?

"Now, once in a while, in such a group of relatives was to be found a man of supreme influence both in his own family group and in the entire tribe; one who was brave and generous and had a commanding personality, the kind that men instinctively follow and obey. (A weak, flabby character did not do.)

"Such a man waited tolerantly while the young and impulsive ones had their say. And then he spoke.

"'My kinsmen, a great wrong has been done to us; we have been outraged and caused to weep, men though we are. . . So of course it would be right for us to go out and do to the murderer what he has done to our loved one. . .'

"He waited. Then he spoke again: 'But, my kinsmen, there is an even better way. That

[2] The trial by ordeal in which a horse was used might be more recent than the Hurdles. The Dakotas did not get horses until sometime in the latter part of the eighteenth century, it is said.

neither his relatives nor ours may continue to feel anger and hate for each other, for we are Dakotas all, we will choose that better way.

"'Go home now, and bring here the best thing you own. It may be a horse, or arms, or apparel, or a robe. They shall be a token of our purpose, for we are going to take this man who has wronged us, and he shall become "something-to-us" in our loved one's stead. Was the dead your brother? Then this man shall be your brother. Or your son? Or your cousin? As for me, he was my nephew, so this man shall be my nephew in his stead. We will give those gifts to him, and he shall be as though he were our loved one returned to us.'

"Whatever the kinsmen felt as he talked, they gradually accepted their leader's proposal for they could see that it was good. This was even better than to retaliate in kind. Shooting would take only an instant and it would be simple, but it would not end the hate. But to win the murderer by kinship, and kindness, was the surer way. It was hard, too, but they undertook to do it. How proud must their murdered kinsman have been, looking backward from Ghostland and seeing what they were doing!

"The council was then informed, and on the appointed day the leading men of the tribe, and the relatives, assembled in the council tipi; and the people, knowing something very rare was about to take place, thronged outside to see and hear as much as possible. The murderer was escorted there. He had not been hiding nor pleading insanity. He was in his right mind; he had done this thing, and he was willing to pay. Whatever the punishment decided for him, he was ready. Only he did not want pity, so he hardened his countenance and did not look about him. He did not want men to say after him in recalling the day, 'Poor thing, how I pitied him! Like some hunted animal he furtively tried to read mercy in men's eyes!'

"But soon the speaker was announcing something. He was saying, 'Smoke now, with these your new kinsmen, for they have decided to take you to themselves in place of One-who-is-not-here. [A euphemism for Dead.] Those goods stacked yonder are for you. Take them, for a token of their purpose to regard you always as they regarded him. Whatever it was that provoked you to such anger, forget it henceforth. And live happy, going in and out among your relatives in perfect confidence that you have their love and loyalty forever!'

"And then the pipe was lighted, and the murderer smoked with the kinsmen of his victim, and so he became something-to-them, that is to say, a relative; and all anger and hate were purged away from all their hearts. Such a man it is said made an even truer relative than many who were related through blood—because he had been bought at so great a cost."

This rare custom seems not to have come up for recording before. After Antelope's account, several Tetons were questioned about it, and some said they had heard of such a thing and others said they had not. Antelope rather thought it was a Yankton tradition, and of Yankton origin, as far as the Dakotas were concerned. I am not prepared to say at this time.

It is not inconsistent with the plains custom of adopting captives. Some Dakotas even adopted enemy warriors who were known to have killed their own relatives. In such cases, the relationship was maintained at a distance, and gifts were sent back and forth at intervals, and when there was a time of truce they visited each other, showing each other every courtesy.

And I have a further sidelight on this, in the instance of a Dakota woman of the Teton group who came upon a slain Crow scout where he lay. He was uncommonly handsome and well dressed. And this woman wailed over him as for her own son. "O, my handsome one, my son! Why have you ventured here, once too often? O, my son!" So she wailed, while she stroked the dead man's cheek and shoulder tenderly. It was known later that her own son, just as handsome, had been killed in a battle with the Crow; and so she had for the moment adopted this man in his place.

The Iroquois, of course, adopted their captives generally; and even whole tribes, which they conquered. These they called "Younger Brother tribes," while they were "Elder Brother tribe." This Yankton custom would then be an understandable feature of the whole pattern of Indian behavior, it would seem.

A PHOTOGRAPHIC STUDY OF THE PROPER MOTION OF NEARBY STARS

PETER VAN DE KAMP

Professor of Astronomy and Director, Sproul Observatory, Swarthmore College

(*Read April 21, 1944*)

It was not until 1718, i. e., one hundred and seventy-five years after Copernicus, that stellar proper motions were discovered by Halley; a century and a quarter more had to elapse before the first cases of non-uniform proper motion were established. This year of 1944 marks the centenary of Bessel's discovery of the variable proper motion of Sirius and Procyon. His interpretation of the perturbing action by close unseen companions was brillantly confirmed through the discovery in 1862 of the companion of Sirius by Alvan G. Clark while testing the 18-inch refractor now at the Dearborn Observatory; the very faint companion of Procyon was not seen until 1896 by Schaeberle with the Lick 36-inch refractor. Later astrometric discoveries of unseen companions were made from visual micrometric measures in two binary systems (ζ Cancri by Seeliger, ξ Ursae Majoris by Nörlund). The potentialities of the accurate photographic method, especially if applied by long focus visual refractors, were gradually realized through the astrometric foundations laid by Schlesinger and Hertzsprung in the fields of parallax and double stars respectively. Unseen companions in several nearby double stars were discovered by Strand. For single stars occasional discoveries have resulted as a by-product of work on parallax, primarily by Reuyl at the McCormick Observatory, where in a limited way a systematic program had been introduced. For the problem of unseen companions, where spectroscopic studies already have so greatly contributed, the systematic astrometric study of orbital motion in nearby faint stars should be very promising and not postponed any longer.

Since 1912 the Sproul Observatory had been engaged in photographic parallax and mass determinations of double stars. After 1937 the emphasis was laid on the present astrometric study which includes all single stars nearer than 33 lightyears that are within reach of the 24-inch refractor. The instrument is particularly suited for this problem because of its long focal length of 36 feet, which gives a scale of $1'' = .053$ mm. With modern fast panchromatic plates ($5'' \times 7''$) and a yellow filter, the visual objective of 24 inches aperture insures a rapid accumulation of accurate photographic positions, referred in each case to a set of three or four faint reference stars of small parallax. The photovisual arrangement reduces atmospheric dispersion to a minimum; wherever needed, magnitude compensation is effected through a rotating disk. A careful study has been made of the various sources of errors; emulsion shift errors are effectively reduced by the use of reversed exposures on the same plate. The principal limitations in accuracy are represented by a probable plate error of $''.027$ for a good plate with two exposures, and by a probable night error of $''.012$ affecting all exposures taken on any one night. The maximum number of exposures is four on one plate, while as a rule not more than two double plates are taken on any one night.

The problem of separating the barycentric proper motion, the annual parallactic motion, and the star's orbital motion presents certain complications. One must watch for spurious and multiple periods made possible by the limited accuracy and by gaps in the observational series. Because of hour angle restrictions not more than six months' observations are available each year covering the retrograde half of the parallactic orbit around opposition. Hence annually recurring gaps of six months exist in any case; these are troublesome when the star's period is close to one year. Of the parallactic orbit all elements are known except the size; of the star's peculiar orbit all elements have to be found.

Except for very long periods, the separation of the components generally would be well below the size of the photographic image ($2''$ to $3''$) and the observed orbit would refer to the blended images of primary and companion. The resulting "photocentric" orbit is assumed to be similar to the relative Keplerian orbit of the two components on the scale $B - \beta$, where B is the fractional mass of the companion and β its frac-

tional luminosity. The approximate validity of this assumption has been confirmed for known close visual binaries.

For several stars the measurements made so far have established a range in deviations from uniform proper motion well exceeding the expected range in errors of observation. A report is made here on the two of the nearest stars for which not only the variability is beyond any doubt but for which a provisional orbital analysis could be made. The objects are the well-known red dwarfs Lalande 21185 (in Ursa Major) and Barnard's proper motion star (in Ophiuchus). For each star the measured positions were reduced to the same origin, scale, and orientation and were then represented by a geocentric ephemeris based on least squares solutions for parallax and proper motion. The residuals were grouped into normal points, averaging 4.3 and 6.3 nights, 15 and 27 plates, respectively; the time range of these points is seldom over two weeks. In each case the behavior of the residuals proved to have a cyclic character; after adjusting for proper motion and parallax, periods of 1.28 and 1.07 years were derived. The normals were grouped according to the orbital phases, and supernormal points formed, 5 and 9 in number. The pattern of the latter could be satisfactorily represented by Keplerian motion which was considered the basis of graphic solutions for the orbital elements. The deviations for the normal points after introduction of the orbit agree closely with the expected error range. Although the Lalande star (208 plates on 60 nights) has only one-fourth as much observational material as Barnard's star (717 plates on 172 nights), its orbit is probably not less accurate since the different orbital phases are more uniformly covered. For both stars the orbits are highly inclined; the eccentricity for the Lalande star is high (.72), that for Barnard's star low (.20). The semi-axis major is ″.054 or .132 astronomical units for Lalande 21185, and ″.06 or .118 a.u. for Barnard's star. In either case the total range of the apparent photocentric orbit proves to be slightly over 0″.1.

The dynamical interpretation of the orbital data leads to a lower limit for the mass of the companion:

$$(B - \beta)(M + m) = \alpha P^{-\frac{1}{3}}(M + m)^{\frac{1}{3}}.$$

Here α is the semi-axis major of the photocentric orbit expressed in astronomical units, P the period of revolution in years, M and m the mass of primary and companion in terms of the sun's mass. The orbital constant $\alpha P^{-\frac{1}{3}}$ is found to be the same (.112) for both objects. Although approximate equality would not be unexpected, the observed identity must be considered an accidental result of the present state of the observational data.

The lower limit for the companion mass in either star is thus:

$$.112(M + m)^{\frac{1}{3}}.$$

As long as the companion remains unseen, astrometrically or spectroscopically, the total mass remains unknown. From the known red dwarf character of the primaries we assume the total mass to be .4 times the sun's mass, which leads to a minimum fractional mass .15 for the companion, corresponding to a value .06 times the sun's mass. Since this is a lower limit, the companion objects undoubtedly are stars. Here the analysis has to stop because of our ignorance of the luminosity of the companions; if the latter would be 3, 2, or 1 magnitudes fainter than the primary, the above fractional mass of .15 would have to be corrected by .06, .14, and .28 respectively! These corrections lead to acceptable results; their actual amounts could be evaluated from spectroscopic studies, since the spectroscopic orbit of the primary probably would not be affected by blending effects. For the range in radial velocity are predicted lower limits of 9 km./sec. for Lalande 21185 and 6 km./sec. for Barnard's star, appreciable quantities thanks to the high inclinations. The visual separation at greatest elongation should be about ″.3 to ″.4 for Barnard's star; because of the high eccentricity and longer period, it may be over ″.5 for Lalande 21185. No elongation of the photographic exposures has been observed.

Whether the companion stars eventually turn out to be red dwarfs or white dwarfs, the present study of individual nearby objects confirms the high state of duplicity indicated by earlier statistical studies (Kuiper). Excluding the sun, the nearest star, α Centauri (distance 4.3 lightyears), contains three component stars. The second nearest star, Barnard's star (6.4 lightyears), is now found double. Wolf 359 is too faint to have been subjected to any proper scrutiny, astrometrically or spectroscopically; Lalande 21185 (8.0 lightyears), vying with Wolf 359 for the third nearest place, is now found

double. Then comes the well-known binary Sirius (8.6 lightyears). Excluding the sun, the five nearest stars thus average two component stars. It seems appropriate to replace the usual formulation of the study of the frequency of double stars by its contrasting aspect, namely, that of establishing the upper limit for the possibly minute frequency of single stars. Quoting Bessel (*Monthly Notices* **6**: 136, 1844): "For, even if a change in motion can, up to the present time, be proved only in *two* cases, yet will all other cases be rendered thereby liable to *suspicion;* and it will be equally difficult, by observations, to free other proper motions from the suspicion of change . . ." For more distant "pseudo-single" objects, however, it will generally be increasingly more difficult to detect by the astrometric method possible deviations of photocentric from barycentric positions.

We conclude by noting that our sun seems a well-established specimen of single star. Although we should refrain from generalizing from the present limited data, the observations suggest a particular status for the sun other than its having planetary companions, which for a long time will be difficult to establish generally for other stars. Rather does it occur to us that from the observational point of view the status of the solar system among the stars may turn out exceptional because it contains only one star. Borrowing from chemistry: the sun appears like an "atom" among its neighboring star "molecules." Here we should like to quote Campbell and Curtis, who in the text accompanying the First Catalogue of Spectroscopic Binaries (*Lick Observatory Bulletin* **3**: 145, 1905), state: ". . . In fact, the star which seems not to be attended by dark companions may be the rare exception. There is the further possibility that the stars attended by massive companions, rather than by small planets, are in a decided majority; suggesting, at least, that our solar system may prove to be an extreme type of system, rather than a common or average type . . ." This possibly exceptional event of the sun having *no* companion *star* would be of particular interest for our notions of the origin and evolution of the solar system; it would point strongly to Russell's theory, which requires a past, but now liquidated, stellar companion of the sun.

As to the newly discovered astrometric binaries, the periods, shorter than for any known visual binary, need not cause surprise; had these been rather larger the duplicity might easily have been discovered visually, unless the components were intrinsically faint (*cf.* Sirius and Procyon).

The present study is a co-operative one, in which over a period of years many individuals have contributed, at the telescope, at the measuring machine, and in various other phases. For both stars the measurements were started by Miss Marion E. Wolff. For Lalande 21185 measurements were also made by Dr. Gustav Land, who in 1941 discovered the variable proper motion for this star; for Barnard's star the variability was noticed as early as 1939. For the past two years measurements were made on both objects by Miss Sarah L. Lippincott. The orbits were computed by Mr. Günther Reuning. Particular credit is due Mr. Roy W. Delaplaine, observer at the Sproul Observatory, for his successful efforts, especially during the last few years, in securing extensive series of photographs on these stars.

REFERENCES

KUIPER, G. P. 1942. The nearest stars. *Astrophys. Jour.* **95**: 209.

RUSSELL, HENRY NORRIS. 1935. The solar system and its origin. Macmillan.

STRAND, K. AA. 1943. The orbital motion of 61 Cygni. *Proc. Amer. Philos. Soc.* **86**: 364–367.

—— 1944. The astrometric study of unseen companions in double stars. *Astron. Jour.* **51** (1149): 12–13.

VAN DE KAMP, PETER. 1944. The astrometric study of unseen companions in single stars. *Astron. Jour.* **51** (1149): 7–11.

POSTNATAL DEVELOPMENT OF THE HUMAN EXTREMITIES

CHARLES BENEDICT DAVENPORT†

TABLE OF CONTENTS

PAGE

Abstract 375
Preface 376
I. Introduction 377
Material 377
Nomenclature 378
Methods of measuring 378
Abbreviations 378
II. Growth in length of limbs and segments 379
1. Gross upper limb 379
2. Net upper limb (without hand) 380
3. "Arm" (humerus) 383
4. Forearm (radius) 385
5. Gross lower limb 386
6. Net lower limb 389
7. Thigh (femur) 389
8. Shank (tibia) 391
III. Change with age of percentage ratios and indices of linear dimensions 392
1. Gross upper limb relative to stature 392
2. Gross upper limb relative to sitting height 394
3. Gross upper limb relative to gross lower limb (intermembral index) 395
4. Net upper limb relative to stature 396
5. Net upper limb relative to net lower limb (net intermembral index) 396
6. "Arm" length relative to gross upper limb length 398
7. "Arm" length relative to net upper limb length 399
8. "Arm" length relative to thigh length (femoro-humeral index) 399
9. Forearm (radius) relative to "arm" (brachial index) 401
10. Forearm (radius) relative to shank (tibia) (radio-tibial index) 404
11. Gross lower limb relative to stature 407
12. Thigh length relative to net lower limb length 408
13. Shank (tibia) relative to thigh (crural index) 409
14. Span relative to stature 412
IV. Change with age of percentage ratios of girths 412
1. Girth of "arm" relative to length of upper limb 412
V. Development of the hand 414
1. Growth of hand in length 414
2. Growth of hand in width 416
3. Percentage ratio of hand width to hand length (hand index) 417

PAGE

VI. Development of the foot 418
1. Growth of foot in length 418
2. Growth of foot in width 419
3. Change with age of foot index 419
VII. Interrelation of certain dimensions 420
1. Interval between height of radial-stylion and tibiale height 420
2. Percentage ratio of gluteal depth to gluteal height (gluteal index) 422
3. Gluteal index relative to stature 424
4. Growth of scapula in length 424
5. Change with age of the angle made by the plane of the scapula with the frontal plane of the body 425
VIII. Discussion 426
1. The reliability of measures of linear dimensions of limbs taken on the living (with W. D. Hallock) 426
2. Growth and differentiation 429
3. Rhythm of growth 432
4. The influence of gravity on dimensions and proportions 437
5. Growth of limbs in relation to sex 437
6. Growth of limbs in relation to social type 438
7. Growth of limbs in relation to race 438
8. Individual limb growth and proportional change 439
9. A comparative study of twins 440
10. Special cases that reveal morphological abnormalities 441
A. Dwarfism 441
a. Cretins 442
b. Chondrodystrophics 442
c. Mongoloid dwarfs 443
d. Other dwarfs 444
B. Acromegaly 445
11. Long bone growth, negative growth, self-sculpturing and responsiveness to stresses of bone 446
12. Man's place among the mammals 449
13. Dimensions and proportions of limbs in relation to function 451
IX. Summary of conclusions 451
X. Bibliography 454

† Dr. Davenport died on February 18, 1944, some time after this monograph had been completed and thoroughly revised by him in preparation for its publication in this form, but before it had been sent to press. His former associate, Dr. Morris Steggerda, has undertaken editorial supervision in seeing it through the press.

ABSTRACT

This study is based on observations of several hundred children between the ages of birth and maturity, considered partly massed and in even greater part individually and "longitudinally." The results are presented graphically in 84 figures. The study is divided into 7 sections, including growth in length of limbs and limb segments, changes with age of percentage ratios and indices of linear dimensions, change with age of percentage ratios of girth, development of the hand and of the foot, interrelation of

certain dimensions, and a discussion of a series of topics, such as reliability of measurements on the living; growth and differentiation; rhythm of growth; influence of gravity on dimensions; growth of limbs in relation to sex, social type, and race; individual limb growth and proportional change; growth in fraternities, twins, and endocrine cases; long bone growth, self sculpturing and responsiveness to stresses; and man's place among the mammals.

The length of the limbs increases with age, but there are juvenile and adolescent spurts. In adolescent girls the limbs may even shorten slightly. In some microsomics growth may continue beyond the ordinary age of cessation of growth. The mean length of limbs differs from that of standard Whites in the Negro race, in acromegalics and dwarfs. The rates of growth in both of monozygotic twins are alike. While the ratio of radius to humerus length diminishes from birth on, that of tibia to femur remains more nearly constant. The excessively long lower limb of Negroes depends more on shank than on thigh. Curves of change of proportions with age differ from growth curves inasmuch as they do not normally slope upwards. Any departure from horizontality is an indication of differentiation in development. Many examples are given. Thus, the upper limb grows more rapidly and the lower limb still more rapidly than sitting height. The lower limbs grow faster than the upper after 2 years; but irregularly before that; in Negroes the difference is not so great. The femoro-humeral and radio-tibial indices gradually diminish after 3 years, departing from the type of brachiating apes with their long upper limbs. The lower limb grows faster than sitting suprasternal height and in acromegalics may become twice as long. Span increases faster than stature, more rapidly in boys than in girls and much more in Negroes than in Whites. Girth of the limbs tends to increase inversely to limb length. The hand index is larger in Whites, smaller in the Chimpanzee and the Negro, and larger in Mongoloid dwarfs. The foot index increases before birth and slowly to childhood, while that of apes diminishes with age. The gluteal masses tend to increase in depth until 8 months post partum. The gluteal index reaches in boys a maximum at about 11 years. The scapulae during early postnatal development tend gradually to assume parallelism with the frontal plane. A criterion for distinguishing between errors in technique and biological changes is developed. The advantages in certain cases of expressing growth in Ångstrom units per second is stressed. Rhythm in growth of the limb bones is wide-spread. The difference between the sexes in dimensions and differentiations is pointed out. The human races differ in proportions of limbs. The growth of the child's limbs does not follow strictly the "law of relative growth." Fraternities may show curves with family resemblance, but in monozygotic twins identity is frequently approached. A comparison of different types of dwarfs—cretins, chondrodystrophics, Mongoloids, and microsomatics—is made. The acromegalic has differential characters that contrast with those of chondrodystrophics. The growing long bones are very responsive to mechanical stimuli both internal and external and both in general form and in structure. The near identity in form of hand bones in monozygotic twins measures the control of hereditary factors. Functionally, man has passed from the group of brachiating animals and come to join the groups of runners, jumpers, swimmers, and even burrowers. He has evolved to a point where he is better fitted for land and water than for the trees. In his ontogeny he shows, at every step, evidence of the path of his evolution.

PREFACE

As a mammal man has become differentiated from the quadrupeds by the more specialized use of his extremities. The cephalad pair is used chiefly for prehension as it is in part in anthropoid apes also and indeed in many lower mammals such as squirrels. The caudad pair is used for plantigrade walking as in the anthropoids, many other primates, many carnivores and insectivores. But, though specialized, the human extremities are all fitted for a variety of functions such as swimming, running, jumping. The present study traces the development of the human extremities in later ontogeny and takes occasion to note certain evidences which the process reveals of man's phylogenetic development.

This study has been made possible by the collaboration of many organizations and those in charge of them (as listed in my 1939 and 1940 papers) and by the statistical assistance of Miss Catherine Carley, Mr. William Drager, and Mr. Wilfred D. Hallock, and the continued literary assistance of Mrs. Inger M. Andersen. Miss Olive Renfroe, of Letchworth Village, took annually repeated roentgenograms of the hands of about one hundred of the children studied. To all of these grateful acknowledgment is made. Nor do I forget to thank the children themselves, who after having reached the age of twelve months co-operated in so friendly a manner.

The expense of the work has been met chiefly by the Carnegie Institution of Washington, in part by Columbia University and Letchworth Village; also by grants made by the American Philosophical Society from its Penrose Fund, by the Dazian Foundation, by the Trustees of the Bache Fund of the National Academy of Sciences, by vote of the Committee of Human Heredity of the National Research Council, and by the American Academy of Arts and Sciences. Only by means of such grants has it been possible to complete this work. This assistance is gratefully acknowledged.

In the work of transcribing data from the original records to statistical sheets, much assistance was received from the Works Progress Administration Project No. 54536 O.P.65–1–21–551, for which the Town Board of Oyster Bay through Supervisor Harry Tappen stood sponsor, while

the Long Island Biological Association afforded special facilities. Acknowledgment is made of assistance from the National Youth Administration furnished through Miss B. Ziegler, Supervisor.

Finally, it should be stated that the paper here offered is only part of the original manuscript. The latter proved to be too long and expensive to publish, and hence this abbreviation. This may help to account for the fact that some "summaries" cover more ground than the text they are supposed to summarize. It seems, however, best to include the entire summary, even though all the evidence for it could not be presented. All the original figures are being preserved at the Genetics Record Office, Cold Spring Harbor, N. Y., and will doubtless be available for study upon application.

I. INTRODUCTION

MATERIAL

In this study both "longitudinal" and mass measurements are utilized. Included are first "longitudinal" [1] series on 34 normal babies from a day or two after birth to 2 or 3 years postnatal. These were measured at the Columbia University Child Development Clinic, Babies Hospital, New York City, the director of which, Dr. Myrtle McGraw, provided fine facilities for work. It is to be understood that the frequencies under one week and over 3:0 years were somewhat less than 34. The babies were measured at either two- or four-week intervals. To fill the gap between 3 and 6 years, 24 boys and 30 girls of a children's home were measured from one to three times each. The above were entirely normal children, 95 per cent of "North European" extraction. To complete this control series, a second large group of children from the Orphan Society of Brooklyn (B. O. A.) were measured "longitudinally." These were United States born, from Protestant homes, and attended public grammar and high schools. They constituted a satisfactory homogeneous lot, well and uniformly fed and cared for. Practically all were of English and Scotch ancestry certainly for one or two generations back. Several families of siblings were included. Of these children, 78 were boys and 64 girls. They ranged from 6 to 18 years. Some were measured only once, but most of them five times. The number of each age observed between 6 and 16 years ranged from 6 to 20 at each half-year age period. These frequencies are not large, but the homogeneity of the group measured made large masses unnecessary. On account of the fact that most of the children were in "longitudinal" series the ordinary computation of standard errors was not applicable. The smoothness of the mean advance of a dimension in the half-year classes was evidence of adequacy of the material. It may be added that the children had not all lost both their parents by death. In many cases the parents were separated or divorced, or one parent had died or was more or less permanently invalided. This series was used as a control in the present study. It seemed suitable for comparison with Mediterranean and Jewish and Negro groups, with the feeble-minded of various grades, and with special endocrine cases.

The third large group of children was studied at Letchworth Village, an institution for the feeble-minded. Three series were used here; one of low-grade (idiot) boys who showed no neurological symptoms. These ranged from 7 to 20 years and over. They were measured half-yearly, some of them repeatedly up to 13 times. There were 130 boys in this series, and they were of various "racial" groups. Some were of old New York State stocks from rural counties, a considerable number were Italians and Jews from New York City, and a few were of German, Scandinavian, and Irish origin. The classification by nationality was possible owing to the reports of social workers who collected such data on their visits to the homes of the inmates.

In the statistical treatment of these children they were grouped as old United States stock and Nordic, Italian, and Jewish. Individuals not falling into these classes were eliminated.

A second series included high-grade feeble-minded boys—the D1 series. They numbered 119 altogether and were measured semi-annually between 7 and 20 years. The frequency varied at the different ages. Thus of the Nordic group the frequency at each half year ranged between 8:6 and 18:6 from 10 to 29 for each age group with a maximum of 29 at 12:6–14:6 years. Of the Italians the frequencies were 5 to 8 at each half year from 9 to 17 years. Of the Jews 9 children were measured repeatedly 3 to 5 at each year-age group from 9 to 15 years. There were also 7 colored boys of $\frac{1}{2}$ to $\frac{3}{4}$ dark

[1] Individual measurements at different chronological ages.

skin color, mostly 3 to 8 at each age between 7 and 15 years.

A third series included high-grade feeble-minded boys and girls. This is the D2 series. Originally planned to include 50 of each sex, the number rose to over 60. In statistical treatment they were classified as in the D1 series. Of the United States and Nordic boys there were about 15 to 28 at ages 9 to 16; of the girls, 13 to 31 at those ages. The frequency of the Italian boys ranged from 10 to 16 at ages 11 to 15; of the girls, from 5 to 8. The Jewish boys ranged from 6 to 9 and the girls from 2 to 5 at the above ages. The numbers as thus broken down are indeed small, but the more homogeneous the group, the fewer individuals are required for significant means. Again these measurements are all from longitudinal series. Some statistical smoothing was done, especially on the smaller series.

The Mongoloid series comprised only a few of the 511 that were measured in different institutions of different sex, age, and racial stock. Only males of United States and Nordic stock measured at Letchworth Village were included with a frequency between 4 and 10 for the different ages from 11 to 20+ years. The inclusion of the means for comparison with other "social groups" seemed indicated. Their growth curves were placed in the social group rather than in the group of special cases, since in the latter group only individuals and not means are considered.

The twins and special cases were studied merely as individuals. The discussion of these cases did not call for means.

Finally, it is important to note that these measurements and observations, especially at Letchworth Village, were taken with very few exceptions by myself alone with the aid of various recorders at different times. Any exceptions were made under my eyes by an assistant whose findings were checked by being recorded on schedule sheets that contained earlier records of the same dimension of the same child. Thus the personal equation, which is apt to be large in anthropometric work, was eliminated. The value of the great Harvard growth study is diminished by the circumstance that a considerable number of observers were at different times employed in the work of gathering data.

The collection of data most resembling mine is that of Godin (1903) on 100 boys between 13 and 18 years. In the publications of Godin's work, means only are considered. The paper of Stuart, Hill, and Shaw (1940) gives individual curves of growth of the tibia.

NOMENCLATURE

A word about nomenclature. The "appendages" ("Gliedmassen") as used by mammalogists are called "extremities" by human anatomists. The word limb is in good popular use and on account of brevity, and to avoid ambiguity, is used here. Upper limb = anterior appendage = upper extremity. This upper limb is composed of the "arm," called also upper arm ("Oberarm") by anthropologists; of the forearm, called also lower arm ("Unterarm"); and of the hand. The lower limb is composed of the thigh, the shank, and the foot. Since, as measured by me, the lengths of the "arm" and the forearm are approximately those of the humerus and radius, these terms will be frequently used in speaking of these segments. Also femur and tibia will be used in speaking of the length of the chief segments of the lower limb. The term gross upper limb includes the hand; net upper limb excludes hand. Gross lower limb includes, net lower limb excludes, the foot.

METHODS OF MEASURING

The babies were measured, as described in my 1938 paper (p. 275), lying down, until at about 13–15 months they were able to stand alone. Sitting height was obtained in the young babies by raising the lower limbs and measuring from head to buttocks; in the older babies by seating them on the table. All older children were measured in a vertical position, and the measurements were taken on the skin. Hermann-Rickenbach instruments were used.

ABBREVIATIONS

In the pages and descriptions of text figures the following abbreviations are used: B. O. A., Brooklyn Orphan Asylum; LVD 1, LVD 2 or D1, D2, Letchworth Village Developmental series; I, idiot series; M, Mongoloid dwarfs; p. a., per annum; p. p., post partum; f., female; m., male; 5:7 signifies 5 years and 7 months. Davenport (1939) (1940) or merely "1939" or "1940" refer to papers of mine on "Post-natal Development of Outer Nose" and "Post-natal Development of the Head," respectively.

II. GROWTH IN LENGTH OF LIMBS AND SEGMENTS

1. GROSS UPPER LIMB

The length of the upper limb was determined in two or three ways: first, by measuring the height from the floor (subject standing erect) of acromion, radiale, and stylion (and sometimes the dactylion) and subtracting the shorter distance from the longer. When dactylion was not measured from floor, to acromion height minus stylion height was added length of hand as determined from a tracing. Secondly, the distance from acromion to stylion was measured directly by anthropometric calipers. The second method yields results 2 to 5 mm. shorter than the first.

The upper limb, as shown in figure 1, attains a mean length of 2.5 cm. at 12 weeks (Scammon and Calkins, 1929: 216)[2] and of about 20.7 cm.

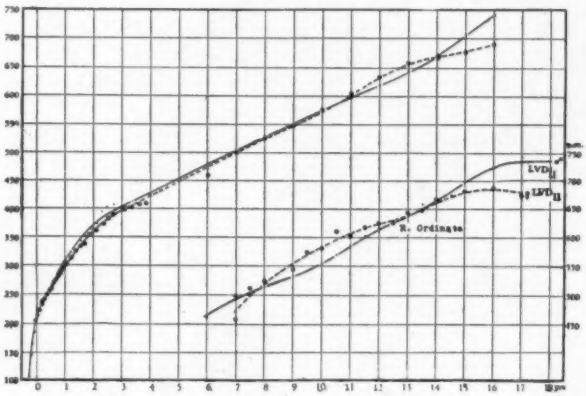


FIG. 1. Standard mass curves of growth of gross upper limb in length, embryo to maturity, U. S. white, male and female, based on fetuses by Scammon and Calkins, Babies Hospital and B. O. A. Prenatal, dots; babies, males, dots; females, circles; B. O. A.: males, dots; females, circles; ordinate scale to the left. Also similar mass curves for LVD 2, males and females; ordinate scale to the right.

at birth. Kugler (1932: 473) finds a mean length at birth of 21.8 cm. for boys and 21.4 cm. for girls, while Friedenthal (1914: 92) finds a mean length of 21.6, Weissenberg (1911: 22) of 21.4 for boys and 21.0 for girls, and Schreiber (1928: 682) of 20.1 cm. for boys and 20 cm. for girls. During the next 2 years the mean limb length increases at the rate of 8 cm. p. a. Thereafter growth slows down, limb length increasing in the mean during the next 10 years, 2.5 cm. p. a., and reaching about 80 cm. in a man who is 180 cm. tall.

Sexual.—Comparison of sex curves in figure 1 shows that the mean upper limb is generally about 5 mm. shorter in girls than in boys. But in connection with the adolescent spurt the mean limb of girls comes, at 13 years of age, to exceed that of boys to the amount of 15 mm. An average length of 72.5 cm. is reached by a woman who is 165 cm. tall.

Social.—The mean upper limb of the Mongoloid dwarfs is 5 to 12 cm. shorter than in the control. This is associated with their generally small body size. In fact, the order of limb length runs in order of general mental (and physical) development: standard, D1, I, and M.

Racial.—The length of the upper limb of the Mediterranean group is about 1 cm. shorter on the average than that of the Nordic; while the Negro group varies with age from 2 to 6 cm. longer than the controls, though the mean statures of the Negro and the white boys are closely similar.

Individual youth.—Figure 2 shows the curves of growth of the upper limb of eight boys between 7 and 18 years. The curves slope upward

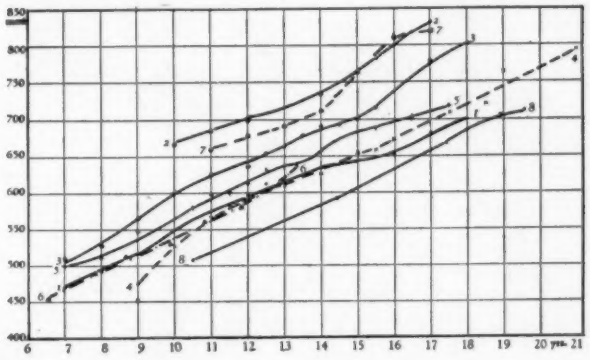


FIG. 2. Individual curves of growth of gross upper limb in length for 8 boys of the LVD 1 series, U. S. (except as otherwise stated), as follows: 1, J. G., no. 114 (Italian); 2, M. G., no. 95 (Italian, acromegalic); 3, E. H., no. 1; 4, C. D., no. 30; 5, C. H., no. 83; 6, J. M., (I ser.) no. 129; 7, R. H., no. 80 (colored); 8, J. Q., no. 156 (Mongoloid dwarf).

with age in similar fashion. Some, like that of the Mongoloid dwarf (no. 8), are approximately rectilinear. Curves nos. 1, 2, 3, 5, and 7 show an increase of slope at adolescence. The uppermost of the curves (no. 2) is that of the acromegalic, and close to that is curve no. 7 of a colored boy.

[2] Friedenthal (1914: 92) finds 4.4 cm. at this age.

Special cases.—In figure 3 the closeness of resemblance between nos. 1 and 2, monozygotic twins, and nos. 3 and 4, dizygotic, is very marked. Naturally the curves of the chondrodystrophic dwarfs (nos. 8, 9, 14) and of the ateliotic dwarfs (represented by nos. 10 and 12) are far below the

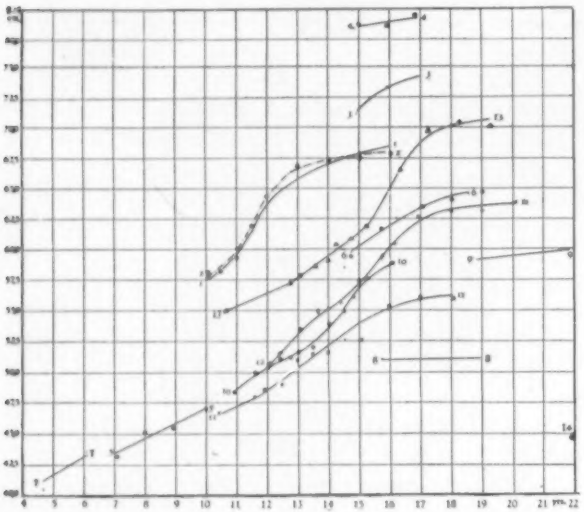


FIG. 3. Individual curves of growth of gross upper limb in length of 2 pairs of twins and 10 special cases: 1, 2, A. G. and M. G. (monozygotic twins); 3, 4, E. D. and S. D. (colored, dizygotic twins); 5, A. A., f. (cretin); 6, C. B., f. (cretin, treated); 7, H. L., m. (cretin); 8, M. C., f. (chondro.); 9, E. S., f. (chondro.); 10, M. C., f. (ateliotic); 11, L. S., f. (ateliotic); 12, A. W., f. (ateliotic); 13, V. R., m. (Italian, microcephalic); 14, W. S., m. (chondro., aged 29 years).

level of the normal twins. The upper limbs of the microcephalic (curve no. 13) are of intermediate length.

Summary.—The upper limb grows in nearly linear fashion from 3 to 12 years, adding about 2.5 cm. in length p. a. Later, at the time of the adolescent spurt, there is usually a temporary increase of slope.

2. NET UPPER LIMB (WITHOUT HAND)

Mass statistics.—Net upper limb length is either the distance from acromion to stylion (the points being measured from the floor) projected; or the sum of "arm" and forearm measured *directly*. The former alone gives the position of the limb segments in relation to the rest of the body. In this section net upper limb length is taken either directly (in babies and the D2 series) or gained indirectly, by subtraction (in the B. O. A. series).

The length of the net upper limb increases rapidly before birth at the mean rate of 17 cm. p. a. (fig. 4). At birth it is about 15.5 cm. Kugler (1932) found 15.4 cm. for boys and 15.1 cm. for girls. During the first year p. p. this rate has slowed down to 5.5 cm. p. a.; during the second year to 4.5; during the third year to 3.5; and during the fourth year to 2 cm. p. a. Thereafter growth continues at about this rate until, during the tenth year and the fifteenth year, in boys, there is an increase, associated with the juvenile and adolescent spurts respectively, to

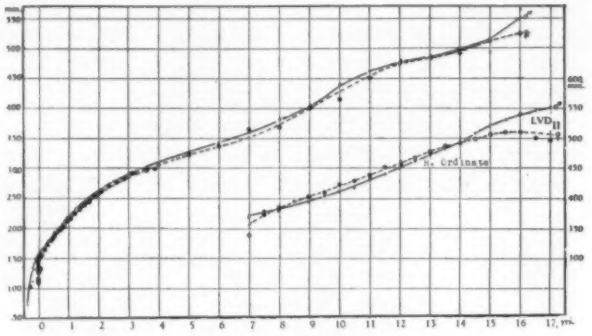


FIG. 4. Standard mass curves of growth of net upper limb in length from midgestation to maturity. Fetuses at Department of Embryology, C. I. of W.; babies at Child Development Clinic, Columbia University; B. O. A., male, entire, and female, broken line; also below at right male and female LVD 2 series all U. S. and North European derivation. Scale of ordinates of D 2 series on right.

about 3.5 cm. p. a. There is an interesting interval between the spurts of growth, when growth is reduced to less than 1 cm. p. a. A length of 55 cm. is attained at 16 years in my control series, and of 60 cm. and more in special cases. Schwerz (1910: 17) finds for Swiss men a mean length of net upper limb of 56.8 cm.

Sexual.—From birth the mean net upper limb length of the girl is less than that of the boy. Among the males in our B. O. A. series the average net upper limb increases in length from 35 cm. at 6 years to 55 cm. at 16 years, or about 60 per cent. In the average female the increases are about the same, but the limb length is about 1 cm. less than in the male between the ages of 6 and 11 years. At the age of the precocious spurt of growth of girls, at about 13 years, the curves of the sexes come close together.

Social.—In the I series the growth curve lies always 1 to 5 cm. below the control series. That of the Mongoloid dwarfs lies 3 to 8 cm. below the

control; this group has extraordinarily short upper limbs, partly in correlation with their short stature.

Racial.—The growth curve of net upper limb length of American Negroes lies 1 to 3 cm. above that of Nordics; of the Mediterranean, about 0.5 to 1.0 cm. below. Thus at 15 years the Negro group has a mean net limb length of 54 cm. This result is, of course, due to the more rapid mean growth of the Negro's arm. Thus from 9 to 15 years the increase is 14 cm. or 2.3 cm. p. a. in the Negroes as contrasted with 10 cm. or 1.7 cm. p. a. in the whites.

Individual babies.—In all cases the curves slope upwards at approximately the same ratio (fig. 5). Yet they do not run parallel. Thus during the first month the net arm lengths of the babies as shown by the curves counting from the shortest to longest is: 4 6 1 7 2 3 5
At 3 : 2 years the order is: 3 6 4 1 5 2 7.
Only one of the three with shortest arms at birth belongs to the group with shortest arms at 3 : 2 years. On the other hand, 2 of those with arm-length above the median at 1 month have arm length above the median at 3 : 2 years. In general, children with long arms at 1 month are more apt to have long arms at 3 : 2 years than are children with short arms to have short arms at 3 : 2 years. Children with long arms at birth tend to retain this superiority; children with short arms at birth are less apt to retain this inferiority. I fear that, since the numbers are so small, these generalizations are valuable only as hypotheses.

Individual youth.—Figure 6 gives the growth curves of net arm length for eight boys. The principal irregularities are associated with the

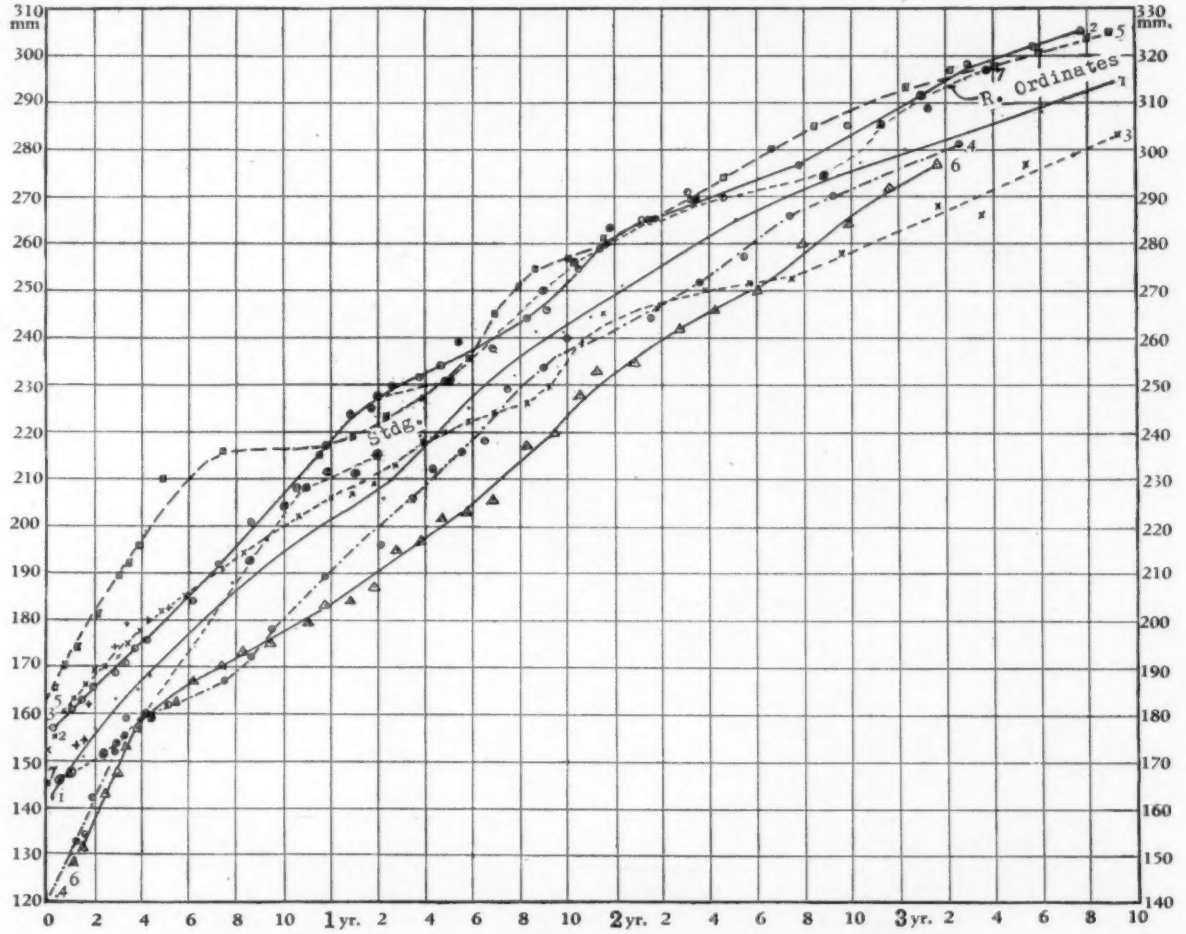


FIG. 5.—Individual curves of growth of net upper limb in length for 7 babies from Babies Hospital: 1, E. K., f.; 2, B. B., f.; 3, C. C., f.; 4, A. L., f.; 5, R. N., f.; 6, F. M., f.; 7, T. P., m. Note that the scale of ordinates for curve no. 7 is on the right.

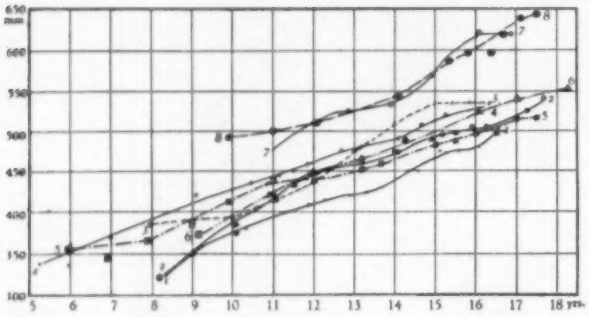


FIG. 6. Individual curves of net upper limb length for 8 boys of LVD 1 series: 1, M. H., no. 2; 2, C. D., no. 30; 3, W. M., no. 8; 4, C. H., no. 83; 5, J. G., no. 114 (Italian); 6, T. V., no. 65; 7, R. H., no. 80 (col.); 8, M. G., no. 95 (acromegalic).

spurts of growth. Curves 7 and 8, at the top, are respectively of an American Negro and the acromegalic.

The question arises here also of the persistence of the relative order of boys from 9 to 16 years. The order from the shortest to longest is at 9 years: 1 2 6 5 3 4 7 8.

At 16 years: 1 5 2 6 4 3 8 7.

In this case all of the children who fall in the short-limbed half at 9 years fall in that half also at 16 years; and those in the long-limbed half at 9 years in that half also at 16 years. The precise order is slightly changed.

Twins and special cases.—Figure 7 gives the growth curves of length of net lower limb for monozygotic twins nos. 1 and 2, 5 and 6, and 7 and 8. These pairs lie close together, in contrast to the curves of the fraternal twins 3 and 4. The ateliotic dwarfs, 11 and 12, like the cretins, have medium long arms. The chondrodystrophic girl of curve 13 and the man at no. 16 have extraordinarily short arms. The microcephalic (curve no. 17) has fairly long arms.

Summary.—The net upper limb length grows parallel to that of the body as a whole with near cessation of growth between spurts. The length of upper limb is less in girls than in boys, very small in Mongoloid dwarfs, large in Negroes. In a group of boys the relative position of limb length remains roughly, but not precisely, the same from childhood to later puberty.

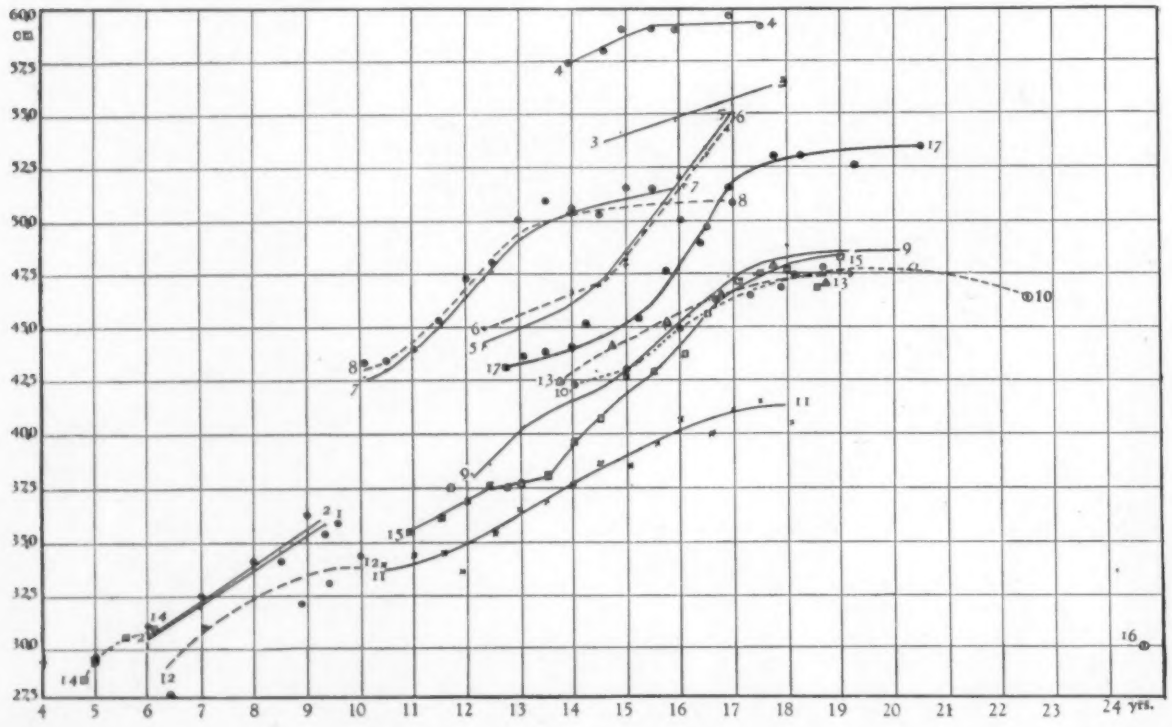


FIG. 7. Individual curves of growth of net upper limb in length of twins and some special cases: 1, 2, D. S. and R. S., mute monozygotic twins, m.; 3, 4, E. D. and S. D., dizygotic twins, f., col.; 5, 6, M. K. and G. K., monozygotic twins, m.; 7, 8, A. G. and M. G., monozygotic twins, f.; 9, 10, A. W. and M. W., microsomic sisters; 11, L. S., microsomic, f.; 12, A. A., cretin, f.; 13, C. B., treated cretin, f.; 14, H. L., cretin, m.; 15, M. C., microsomic, f.; 16, W. S., chondrodystrophic, m.; 17, V. R., microcephalic, m.

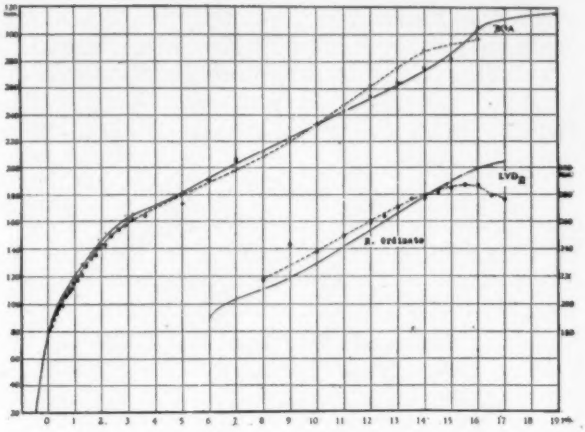


FIG. 8. Standard mass curves of growth of "arm" (humerus) in length. Made up as in fig. 4. Below, D2 series, ordinates at right.

3. "ARM" (HUMERUS)

The proximal segment of the upper limb ("arm" of the anatomists) is in man the most dominating of the three segments: arm, forearm, and hand.

The mean growth curve of length of the "arm" from mid-gestation to maturity is shown in figure 8. The increase is from 1.9 cm. at 15 weeks (Scammon and Calkins, 1929: 218; Friedenthal, 1914: 92) to 8.0 cm. at birth. The growth curve increases 40 mm. the first year, 26 mm. the second, and 16 mm. the third (fig. 9). Thereafter to 13 years it increases 102 mm. or 10.2 mm. p. a. on the average.

Social.—The control boys have the longest humerus, certainly after 12 years.

Racial.—The Mediterraneans after 12 years

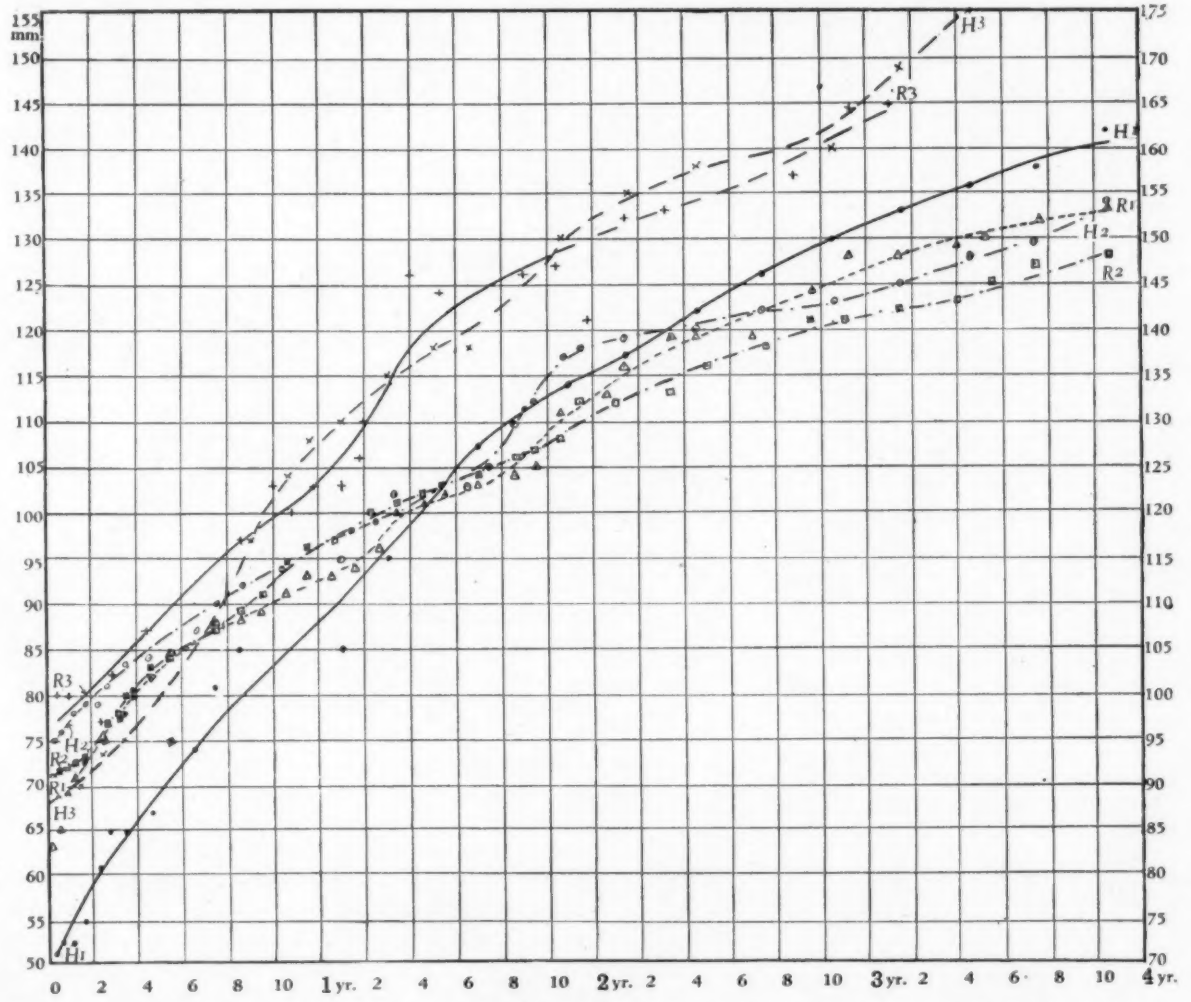


FIG. 9. Curves of growth of 3 babies to compare the growth of humerus *H* and radius *R*: 1, E. K., f.; 2, C. C., f.; 3, T. P., m. Ordinates at left for radius; at right for humerus.

are below the control, and the Negroes are always from 5 to 16 mm. above.

Individual youth.—Figure 10 shows curves of growth of six boys and one girl of the D1 and D2 series. In general, the humerus grows rapidly during the period from 6 to 16 years,

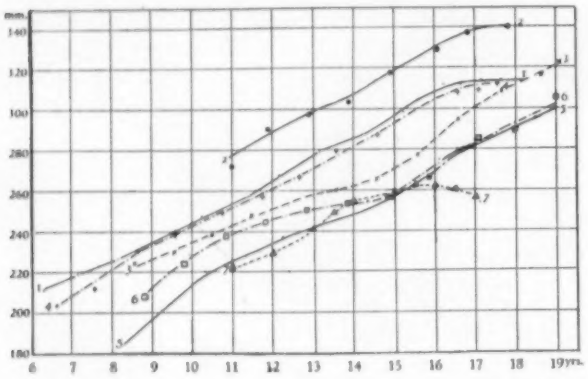


FIG. 10. Curves of growth of "arm" (humerus) in length for 7 individuals of LVD series: 1, D. F., no. 57, m. (Ital.); 2, R. H., no. 80, m. (col.); 3, J. C., no. 10, m.; 4, S. B., no. 55, m.; 5, M. H., no. 2, m.; 6, H. M., no. 25, m.; 7, G. H., D 2, no. 52, f.

at a rate of about 11 mm. p. a. In the case of the girl, no. 7, a period of retarded growth is found between the fourteenth and fifteenth years following the adolescent spurt. A decrease of length occurs in the girl after 16 years. In 3 boys the humerus is still growing fairly rapidly up to 19 years. In all girls where measurements of the arm were continued beyond 15 years, there is a reduction in the length of the humerus (fig. 11). The decrease in length of the arm in the course of one or two years may

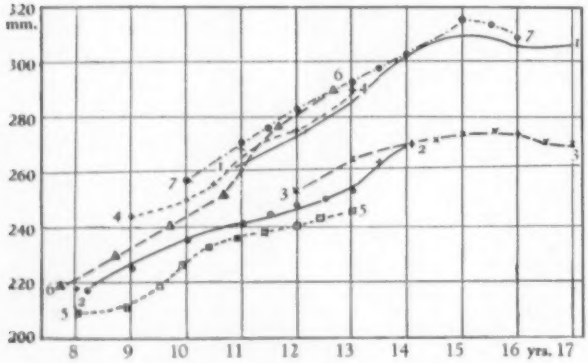


FIG. 11. Curves of growth of "arm" (humerus) in length in 7 girls in the D2 series: 1, E. P., no. 86; 2, G. T., no. 106; 3, A. G., no. 48; 4, R. B., no. 120; 5, O. D., no. 121 (col.); 6, H. B., B. O. A., no. 12; 7, E. A., no. 124.

amount to as much as 12.5 mm. To those who have not followed the growth of long bones in the living, it might seem unreasonable that a growing bone should diminish in absolute length. However, the bone in the living child is a highly plastic thing, and during the period following the rapid growth in length there is, we may guess, in girls at least, a tendency toward increase in diameter, to affect which material is, as it were, drawn from the ends of the bone; or, in other words, the rapidly growing cartilaginous ends of the long bone may come to lie closer together on account of a movement of the material of the diaphysis from the longitudinal to the transverse axis of the bone. By this means the bone is strengthened to an important degree. That this reduction in arm length is a very definite episode is indicated by the fact that in a pair of monozygotic girl twins studied, whose curves are shown in figures 7 and 12, there is a retardation, cessation, or inversion of growth of the humerus in both, beginning at the age of 15 years.

Figure 12 shows first the curves of growth of the length of the arm in the monozygotic twins Angelina and Mary (nos. 1 and 2). It is seen that the points as found and recorded on the graph lie close together, rarely more than 2 mm. apart. One notes that the curve of growth advances rapidly, about 16 mm. p. a. from 10 to 12 years of age, and even more rapidly during the immediate weeks following, but after 13 years growth has been practically completed. Then follows the shortening of the humerus, just referred to. The closeness of the curves in the case of these monozygotic twins is striking when the pair of twins is compared with the other curves on the figure, all with the same ordinates. One notes the great diversity in size of the dimension and in the form of the curves of its growth.

Curves 5 and 6 are of the dizygotic Dot. twins, who are American Negroes. One notes first that the differences in length of the dimension at any age, such as 15 years, amounts to 20 mm. or more. Both curves, however, are advancing at 16 years, at an age when in the Gar. twins the dimension has ceased to increase—is, indeed, diminishing. At 16 years the length of the humerus in the Dot. twins is 30 mm., or more, greater than in the Gar. twins. This is a difference of nearly 12 per cent. It is doubtless due to a difference in the genes that determine directly or indirectly the growth of the humerus,

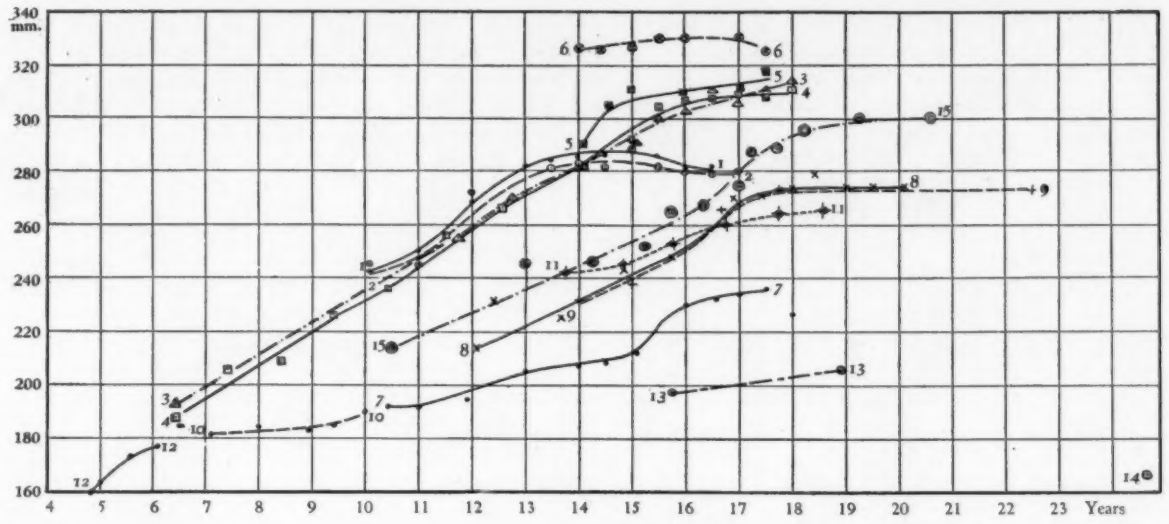


FIG. 12. Growth of "arm" (humerus) in length in twins and some special cases: 1, 2, A. G. and M. G., monoz. twins, f.; 3, 4, Ri. O. and Ro. O., monoz. twins, m.; 5, 6, E. D. and S. D., f., diz. twins (colored); 7, L. S., microsomic, f.; 8, 9, A. W. and M. W., microsomic sisters; 10, A. A., cretin, f.; 11, C. B., treated cretin, f.; 12, H. L., cretin, m.; 13, M. C., ateliotic, f.; 14, W. S., chondro., m.; 15, V. R., microcephalic, m.

that the length in the American Negro is so much greater than that of white children.

No. 10 is the fragmentary growth curve of a young cretin, while no. 11 shows growth of the humerus in a treated cretin at a later age. It was very slow after the seventeenth year. No. 12 in the lower left-hand corner is that of a cretin of hybrid (White-Korean) origin, whose growth was followed, however, during only about one year. This curve shows the very short humerus, namely 160 mm. at 5 years. Growth is rapid at first, but undoubtedly the future development will be as in no. 10.

Curve no. 15 is that of V. R., a microcephalic boy of below control stature. His humerus length in the eleventh year is about 25 mm. below that of the Ols. twins. The length increases less rapidly than in the Ols. twins, namely at 10 mm. p. a., until the sixteenth year, when it increases to 12 mm. p. a. for a time, but at 19 years it begins to approach an asymptote. At the eighteenth year the length of the humerus is about 16 mm. less than that of the Ols. twins. We have noticed elsewhere the especially rapid growth of this microcephalic after puberty has set in. The speed of growth is as though to compensate for a previous retardation in growth.

Summary.—The growth of the "arm" is of interest as approximately that of a single bone (humerus). It reveals very beautifully the variation in rate of growth, decreasing from mid-gestation. In individuals the growth curve is often wavy, and the period of rapid growth in some cases coincides with a slow growth of the forearm. The parallelism of growth of the humerus in twins is striking. Some girls after adolescence show a shortening of the humerus, suggesting a shortening of the bone due to rearrangement or loss of materials. In ateliotic dwarfs the humerus often continues to grow at an age when that of control children has ceased growing.

4. FOREARM (RADIUS)

Forearm length is the projected distance, measured from the floor, of radiale height minus stylion height.

Figure 13 shows the mass curve of growth in length of the radius from fetus to maturity. The advance in fetal life is at the rate of about 11 cm. p. a., which is relatively enormous. It gives some idea of the furious activity of the cells of the cartilaginous zone at the tips of the bone that they should produce collectively in fetal life a zone about 1 cm. thick per month or 0.3 mm. per day (perhaps mostly at night). Eventually a length of 25 cm. is attained.

Of the four social groups, the control B. O. A. series is constantly above the others (mean at 16 years, 24.5 cm.); the Mongoloids strikingly below (20 cm.). The order of all four is that of decreasing intelligence of the groups; and that is, in general, related to body size.

Racial.—Of our three groups, the Mediterraneans have the shortest radii (at 17 years 24 cm.); the Negroes the longest (at 17 years 25.6 cm.), as Todd and Lindala (1928: 73) pointed out. Some African pigmies have radii only 20 cm. long.

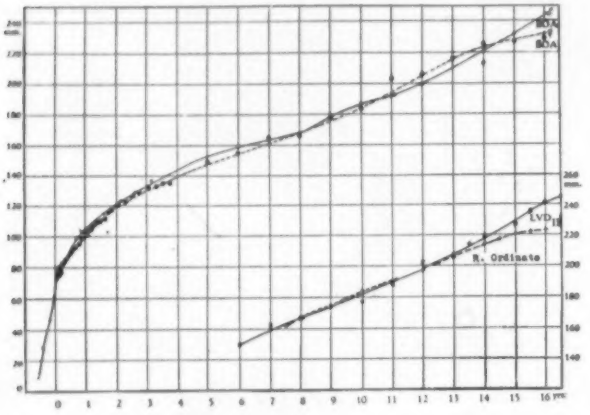


FIG. 13. Standard mass of curves of growth in length of forearm made up as in fig. 4.

Individual babies.—The growth curves of babies, starting at between 7.0 and 8.0 cm. at or near birth, advance to 12.3 to 14.6 cm. respectively, at 3½ years. The radius grows faster in boys than in girls. The slope is not uniform. Some curves come near to forming a straight line. Others show a rapid upward slope at 1:8 years, as growth of humerus slackens (fig. 9). A change of slope that sometimes occurs at about 1 year is probably associated with the change to vertical measurements from horizontal ones at that time.

Individual youth.—Figure 14 gives the length of the forearm in six boys of the D1 series, selected mostly because of their age range. No. 5, of a colored boy, and no. 6, of an acromegalic, are in a class by themselves. All rise gradually from childhood to maturity; usually in wavy, sometimes, as in nos. 3, 4, and 6, in nearly linear fashion, increasing at the average rate of about 7 mm. p. a. In other cases, as in curves nos. 1 and 5, there is a rapid increase in forearm length at the time of the adolescent spurt in growth. The increase in some cases amounts to 20 mm. in the course of 12 months, after which the growth continues at about the average rate again. Children with short forearms do not keep their relative positions after this spurt in growth. This is because of a variation of speed of growth in this group. But the children with long forearms keep their relative positions. For example, the increment is roughly the same for each of the children nos. 4, 2, and 6.

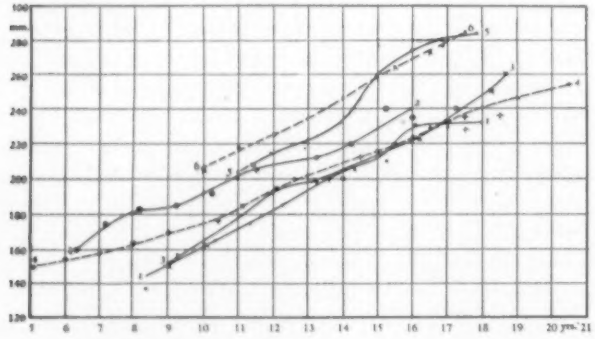


FIG. 14. Individual curves of growth in length of forearm in 6 boys of D1 series: 1, M. H., no. 2; 2, D. F., no. 57 (Italian); 3, C. D., no. 30; 4, C. H., no. 83; 5, R. H., no. 80 (col.); 6, M. G., no. 95 (acromegalic).

Twins and special cases.—The curves in figure 15 of twins 1, 2; 5, 6; 7, 8; show their usual close position. Curves 7 and 8 practically coincide. The chondrodystrophics, 12 and 15, show extraordinarily short radii.

Summary.—The radius grows rapidly before birth, slows up soon thereafter. The Negro radius is exceptionally long. Some boys show a spurt in growth of the radius at adolescence; some do not. There is a striking resemblance in many of the growth curves of radius in single fraternities.

5. GROSS LOWER LIMB

The lower limb is an important part of stature. It is about half of stature in late childhood, but thereafter is usually a larger fraction. Its growth depends on that of the great long bones, femur and tibia, and bones of the foot. The study of its growth in length is largely that of the growth of long bones.

Method and material.—The precise measurement of the lower limb and its bones is, unfortunately, in the living fraught with great difficulties, and many approximations have been proposed. Especially is it hard to find the upper end of the femur, buried as it is in the acetabulum and surrounded by muscle. Some of the substitutes for the upper end are:

(1) Anterior superior iliac spine—*b; b* varies with age, sex, stature; in adolescence it equals on the average 40 mm. If the child is under 130 cm. tall, *b* = 15 mm.; if between 130 and 150, 20 mm.; if between 151 and 165, 30 mm.; 166

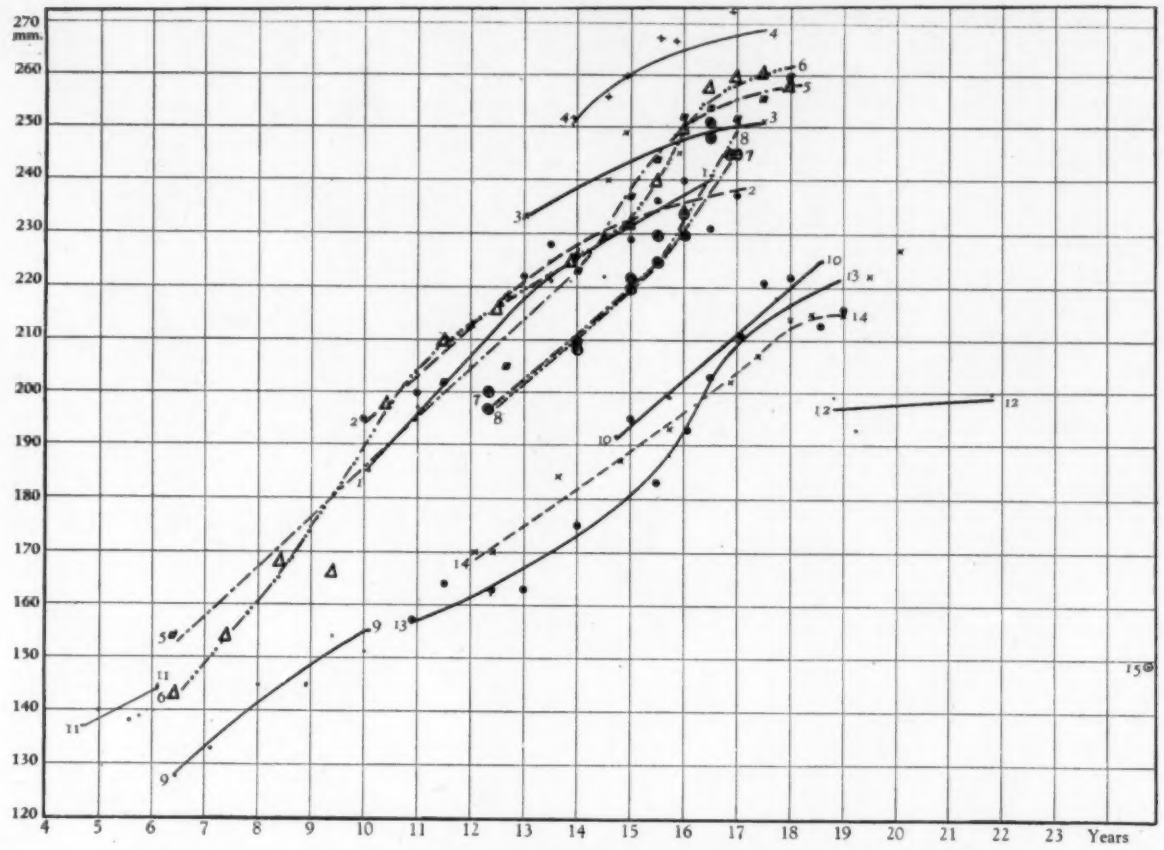


FIG. 15. Individual curves of growth in length of forearm in twins and some special cases: 1, 2, A. G. and M. G., f., monoz. twins; 3, 4, E. D. and S. D., f., diz. twins; 5, 6, Ri. O. and Ro. O., m., monoz. twins; 7, 8, M. K. and G. K., m., monoz. twins; 9, A. A., f., cretin; 10, C. B., treated cretin, f.; 11, H. L., cretin, m.; 12, E. S., chondro., f.; 13, M. C., ateliotic, f.; 14, A. W., ateliotic, f.; 15, W. S., chondro., m. (aged 29 years).

and 175 cm., 40 mm.; over 176 cm., 50 mm. Some authors omit *b*.

(2) Symphysion + 35 mm. on the average.

(3) Trochanter height (very unprecise) + 90 mm.

(4) Perineum height + 90 mm.

(5) Stature minus sitting height (a distant approximation).

(6) In this paper gross lower limb length is obtained, as recommended by Jazuta (1925) and Gray (1941), by taking half the sum of height of the symphysion and the anterior iliac spine. This gives nearly for any age the vertical distances from the floor of the upper surface of the head of the femur. See, e. g., figure 362 of Jackson-Morris, *Human Anatomy*, ninth ed., 1933. It is to be hoped that where feasible this landmark may be used in finding gross leg length.

Control mass curve.—Figure 16 gives the curve of growth of the lower limb from 10 weeks' gestation to past puberty. This curve for the first year is logistic, with inflection at birth and marked slowing down—a concavity above—at

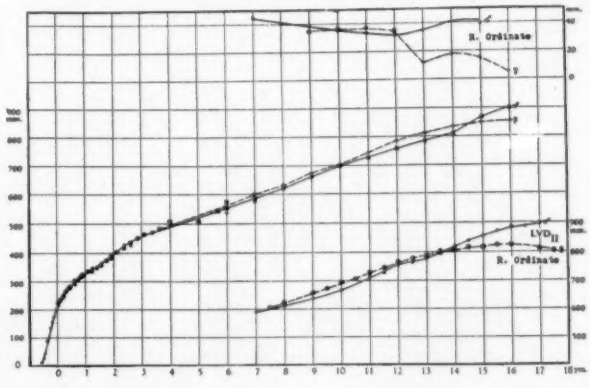


FIG. 16. Standard mass curves of growth in length of gross lower limb made up as in fig. 4. Above, *increments* of B. O. A. curve, male and female; ordinates to the right.

11 to 24 months. This suggests a disturbance due to the assumption of the standing posture. Thereafter (following the unbroken line) growth progresses fairly uniformly with a slowing up at 8 years, a minor maximum at 10 years, and a great increase in slope after 14 years, which represents the adolescent spurt of growth. The curve reaches 90 cm. at 16 years.

Sexual.—The curve of lower limb growth for girls lies above the curve of the boys except possibly at 10 years, when the curves of the sexes nearly meet. From 10 to 12 years the lower limbs of the average girl grow about 12 mm. per year faster than those of the male. After 12 years the growth of the limb in the girl begins to slow up, with approaching maturity. A decussation occurs at 14.5 years.

Social.—The control boys have the longest mean gross lower limb (at age 14 years, 82 cm.). For the idiot group at the same age the mean length is 79.5 cm.; for the Mongoloid dwarfs it is 70.5 cm. The lower limb of the Mongoloids at 7 years is only 2 cm. shorter than standard. It falls behind the standard until at 13 years it is 12.5 cm. below; it then is influenced by the adolescent spurt, but at 16 years has almost reached the asymptote, while the average of the control boys is increasing 4 cm. per year. In

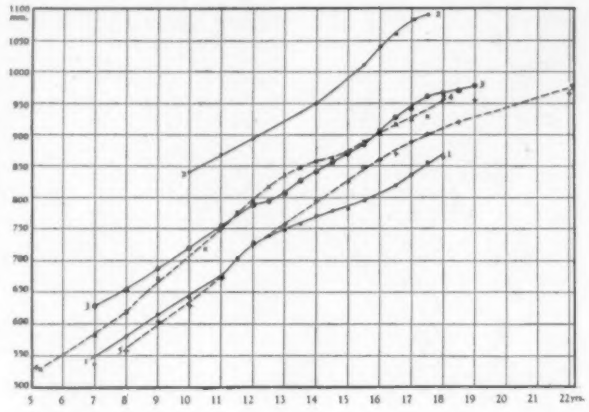


FIG. 17. Individual curves of growth in length of gross lower limb of 5 boys of D1 series: 1, J. G., no. 114 (Italian); 2, M. G., no. 95 (acromegalic); 3, E. H., no. 1; 4, C. H., no. 83; 5, C. D., no. 30.

normal boys epiphyseal union usually occurs at about 20 years. On the average, in these dwarfs the growth of the long bones is retarded—ages 8 to 12 years—and ceases precociously.

Racial.—In our series the lower limbs of the Negroes at 15 years are longest, at 87.8 cm. (*cf.* Todd and Lindala, 1929: 74); Nordic group next, at 82.5 cm.; Mediterranean least, at 77.8 cm. The Negro group advances faster than the

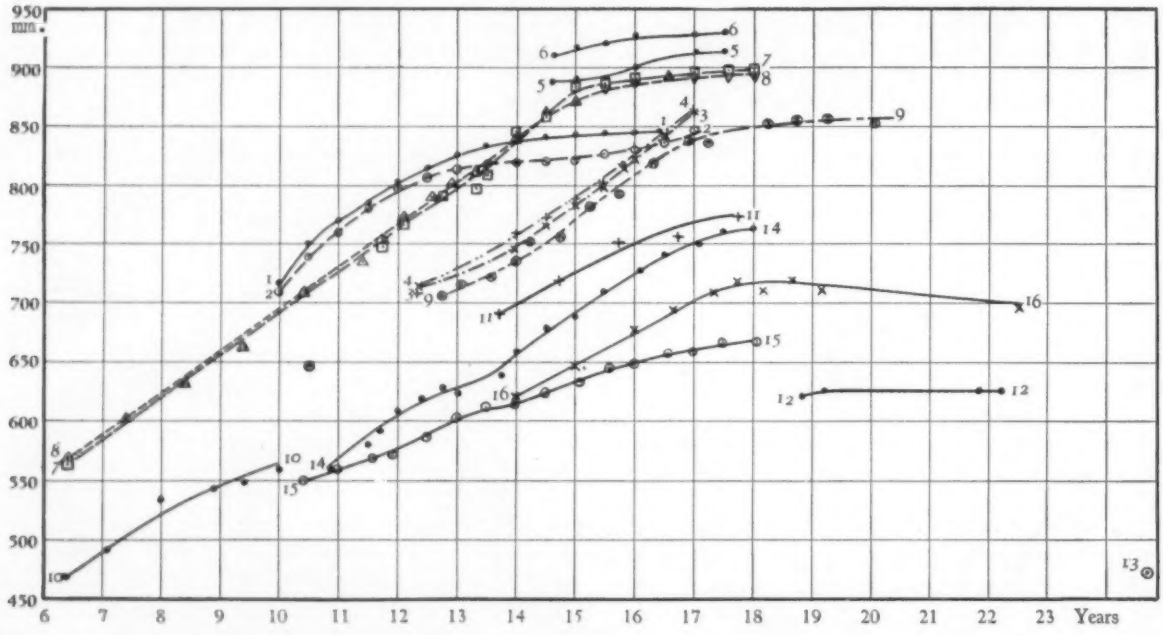


FIG. 18. Individual curves of growth in length of gross lower limb for twins and some special cases: 1, 2, A. G. and M. G., f. (monoz.); 3, 4, M. K. and G. K., m. (monoz.); 5, 6, E. D. and S. D., f. (colored, diz.); 7, 8, Ri. O. and Ro. O., m. (monoz.); 9, V. R., m. (Ital. microcephalic); 10, A. A., f. (cretin); 11, C. B., f. (treated cretin); 12, E. S., f. (chondro.); 13, W. S., m. (chondro., 29 yrs.); 14, 15, 16, M. C., L. S., and M. W. f. (ateliotic).

Nordic, doubtless owing to a greater capacity for long bone growth.

Individual youth.—As figure 17 shows, the lower limbs of these boys elongate steadily with age, but with position and slope differing. The uppermost curve is that of the acromegalic boy whose limbs at 14 years are 95 cm. long (control, 81 cm.). The slope of this curve at 14–15 years is 4 cm. p. a., as compared with 2 to 3 cm. for the other curves on this figure. All but curve 5 show a wavy advance.

Twins and special cases.—(Fig. 18). The development in monozygotic twins is strikingly similar, as curves 1, 2; 3, 4; 7, 8 indicate. It has been difficult to plot these curves as two lines, so often do the values at a given age coincide. The curves of the dizygotic twins, nos. 5 and 6, lie about 2.5 cm. apart. The curve of the untreated cretin, no. 10, lies low in the series, but that of the treated cretin, no. 11, occupies a more nearly median position. The curves of the chondrodystrophics, nos. 12 and 13, occupy low positions. The lower limbs of the chondrodystrophic man (no. 13) have the length of those of a 3-year-old child. They apparently stopped growing at about that age. Perhaps they were proportionately short at birth. The curves of the 3 ateliotic dwarfs (nos. 14, 15, and 16) occupy a low position. The limb length of one girl seems slightly reduced after puberty. All of these measurements are less than the 105.9 cm. leg length found by Henckel (1929: 578) for a 23-year-old, 174-cm.-tall, eunuchoid man.

Summary.—The gross lower limb grows in length in geometric fashion until after birth, slows up in the second year p. p., and then grows steadily with slight increase of slope at 9 and 14 years. The lower limbs of the girls are about 1 to 3 cm. longer than those of boys until $14\frac{1}{2}$ years of age. The curves of average increment of growth show spurts of increments for the male at 8 and 15 years, and for the female at 12 years. These are the ages of the spurt of body growth. Races of mankind differ greatly in lower limb length. The Negroes have the longest limbs, the Mediterraneans the least; the Mongoloid dwarfs have shorter limbs than have the other series. Of individuals, an acromegalic has the longest legs; the Italian boy, the shortest. Babies show a reduction of lower limb length when first measured on their feet. Monozygotic twins have limbs of closely corresponding length at all ages.

6. NET LOWER LIMB

Figure 19 gives the growth changes of net lower limb, control series. Beginning at 12 weeks with a mean value of 2 cm., the limb increases to 20 cm. at birth, or at the rate of 34 cm. p. a. The rate falls in the first year p. p. to 9.5 cm. p. a. and in the second year, to 6 cm.; then to 5.5 cm. in the third and to 4 cm. in the fourth.

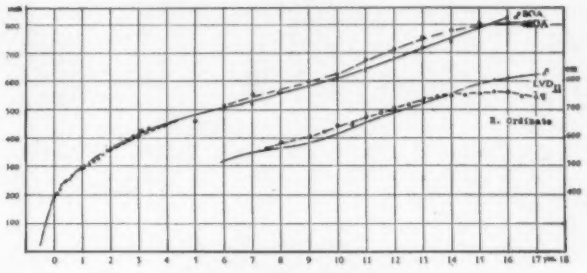


FIG. 19. Standard mass curves of growth in length of net lower limb (without foot) made up as in fig. 4.

In the eighth year the increase is 3 cm. p. a. After 10 years the increase is more rapid, up to 4 cm. p. a.

The net lower limb of adult man is about 90 cm. long. Papillaut (1902: 420) found a mean lower limb length of 79 cm. for adult Parisian males, and Wilder and Pfeiffer (1924: 59) found 78.7 cm. for Smith College women. The length of the net lower limb in anthropoids is (in relation to trunk length) shorter than in man, and in order of reduction from greatest to least as follows: gibbon, gorilla, chimpanzee, orang, man (Mollison, 1910: 124–125). Man's long lower limbs are admirably adapted to swift walking and running, but poorly adapted to life in trees.

7. THIGH (FEMUR)

This dimension was obtained by subtracting tibia height (standing) from gross lower limb length. The long human thigh is a striking trait among the primates. The curve of growth in thigh length of our control series is shown in figure 20.

At the end of the first year the growth in length of the thigh slows down, owing to the measurement of the baby standing, and the consequent shortening of the thigh by gravity. After 2 years growth is proceeding with ever diminishing increments until 6 years of age. Thereupon ensues a rise in slope to a slight maximum at 12 years in the male, and again an

upward slope at 14 years. At 16 years the mean value for the male is 45 cm. and increasing. In the D2 series the curve of the mean male has become nearly asymptotic at 45.0 cm., age 17. In adult German males 180 cm. tall, Bach (1927: 234) found a mean thigh length of 49 cm.

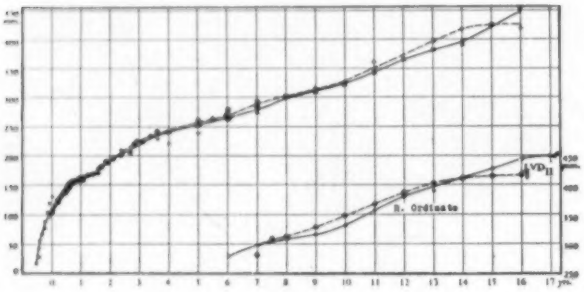


FIG. 20. Standard mass curves of growth in length of thigh made up as in fig. 4.

Social.—Of our four social groups after 11 years, the curve of thigh length of the control series has the highest position, and of the Mongoloid group the lowest. Thus at 16 years the thigh length of mean Mongoloid groups is 36.5 cm., or over 7.5 cm. less than the mean B. O. A. group. However, at 7 years of age the two are only 1 cm. apart. Apparently the Mongoloid is not clearly differentiated from the control at 7 years, but for the next 5 years gains only 0.6 cm. p. a., while at the same time the B. O. A. series is gaining 17.5 mm., or 1.75 cm. p. a. After 12 years the Mongoloid slope increases so that the thigh adds on the average 1.5 cm. p. a. before growth stops at 17 years. That this sudden active growth after a slow period is not merely a mass statistical result is shown by the individual curve of J. Q., whose thigh grew between 12 and 18 years at the rate of 2.2 cm. p. a.

Individual babies.—The curves do not increase in geometric fashion; their increase is of the arithmetical sort. There is, indeed, no reason to expect that as the bone gets longer it will grow proportionally faster, since growth in length is exclusively at the termini. The blood supply of the longer bones may indeed be greater than that of the shorter bones, but the curves as drawn do not indicate that there is any absolute increase in rate of growth of the bone as it increases in length.

Individual youth.—Figure 21 gives the growth of the thigh in length in 7 boys of the D1 series. At 6 years the thigh has increased in length from 250 (in children at 4 years) to 270 mm. Thereafter growth continues more or less steadily to 21 years, reaching a length of over 500 mm. Thus in the course of some 21 years from birth

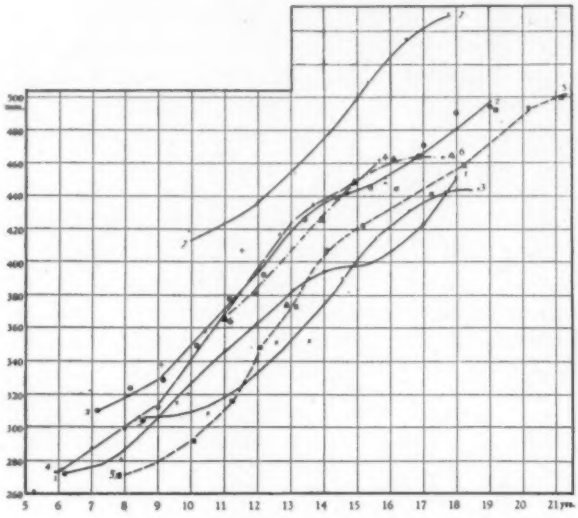


FIG. 21. Individual curves of growth in length of thigh of 7 boys of D 1 series: 1, J. G., no. 114 (Italian); 2, E. H., no. 1; 3, J. C., no. 10; 4, C. H., no. 83; 5, C. D., no. 30; 6, R. H., no. 80 (col.); 7, M. G., no. 95 (acromegalic).

the femur increases in length nearly five times. Growth of the femur in these boys is not, however, uniform. The action of the adolescent spurt, probably due to activity of the pituitary hormone, is very clearly seen in practically all of the curves. It is followed by a slowing down of the growth of the thigh bone. However, curve no. 5 at 20 years indicates that in some boys growth of the thigh is continuing at that age. There are changes of slope probably due to other causes. Noteworthy is the extraordinarily long femur of the acromegalic (curve no. 7).

Summary.—The femur is the largest bone in the body and its growth is strictly terminal. The growth of this bone shows what a long bone can do in this matter. During the last 6 months of gestation it increases 90 mm., or at the rate of 18.1 cm. p. a. or 1.5 cm. per month or half a millimeter a day. The cells of the two cartilaginous ends of the bones are producing that much thickness of collagen at the ends of the columns of cartilage cells besides what goes into the building of the supporting columns. The amount of this material used in the upbuilding of the femur varies with general bodily and

mental development and with race. Mongoloid dwarfs are slow in making a start at rapid growth. In Negroes the terminal chondrogenous processes are more active than in Whites. Terminal growth of the femur is generally linear (additive) after 2 years p. p. At the time of the adolescent spurt, growth is speeded up earlier and more vigorously in Negro children than in whites. The control of growth of femur by genes is well brought out by the similar growth of twins.

8. SHANK (TIBIA)

This is the distance between the tibiale (taken standing, except in young infants) and the terminal tip of the internal malleolus, so far as it could be reached by pressing with the thumb nail through the skin and through the ligaments in which the tip is buried.

The mass control curve of growth of the shank (fig. 22) is in some respects like that of the femur, but its rise is not so fast. The eventual

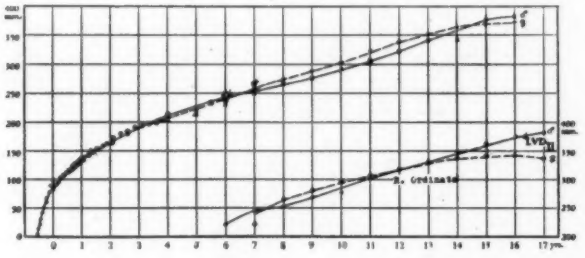


FIG. 22. Standard mass curves of growth in length of shank (tibia) made up as in fig. 4. In the upper curves from the fifth year on the data were taken at the Embryological Laboratory of the Carnegie Institution of Washington.

length of the tibia is 37.5 cm. for the standard males at 16 years, and 38 cm. for 17-year-old boys of the D2 series. Bach (1927: 234) finds for Germans of mean 177 cm. stature a mean tibia length of 39 cm.

Social.—The boys of the control series have the longest tibia after 12 years. After 15 years this order of tibia is: B. O. A., D2, I, and Mongoloid dwarfs. At 16 years the tibial length of the Mongoloids is 33 cm. as contrasted with that of the B. O. A. boys of 39.0 cm.

Racial.—Of our three racial groups, the Negroes have by far the longest tibiae at 15 years (3.7 cm. excess), though their superiority is not great at 9 years, when it is only 0.5 cm. The Nordics occupy a middle position, and the Italians have shortest tibiae.

Individual babies.—Figure 23 gives curves of growth of the shank for four babies, from near birth to the fourth year. The advance is for the most part by fairly regular series of waves, although in the first quarter of the second year there is a cessation of growth in three of the four curves, due to shortening on standing.

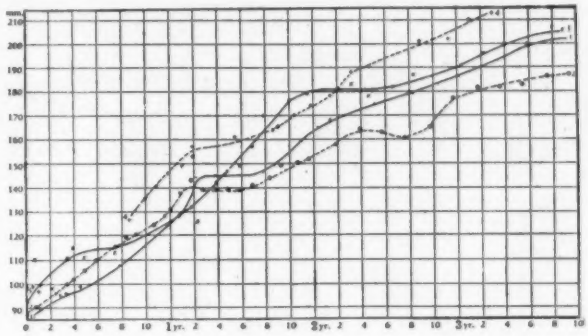


FIG. 23. Individual curves of growth in length of shank (tibia) for 4 babies: 1, E. K., f; 2, C. C., f.; 3, R. M., f.; 4, T. P., m.

The remaining curve (no. 3) shows a cessation of growth at about 7 months and again at 2 years.

Individual youth.—Figure 24 shows the growth of the net lower leg for seven boys of the D1 series and for one boy, no. 8, of the idiot series. At 14 years the range of length is very great—from 32.2 cm. to 41.4 cm. The slope is in most curves steadily upward with occasional accelerations, less marked than in the case of the thigh bone, and most marked at about 13 to 15 years, in connection with the general spurt in total stature. The slope of the curves diminishes after the adolescent spurt, in the case of curves nos. 1, 6, and 8 practically to 0; whereas the curve of the acromegalic boy with the longest thigh, no. 5, is advancing at 16½ years. The length of the tibia attained in the different curves varies roughly proportionally to stature of the boys at 18 years.

Summary.—The mean length of shank is, from fetal life on, less than that of the thigh; and growth is slower, so that a mean length of only 40 cm. is attained as contrasted with a length of 48 cm. of the thigh. The shortness of the shank of Mongoloid dwarfs appears already in childhood. The shank of an adult chondrodystrophic (W. S., m.) is only as long as that of a three-year-old child. Acromegalics have extraordinarily long shanks as well as thighs; and in Negroes the excessive length of the shank is no less marked than that of the thigh.

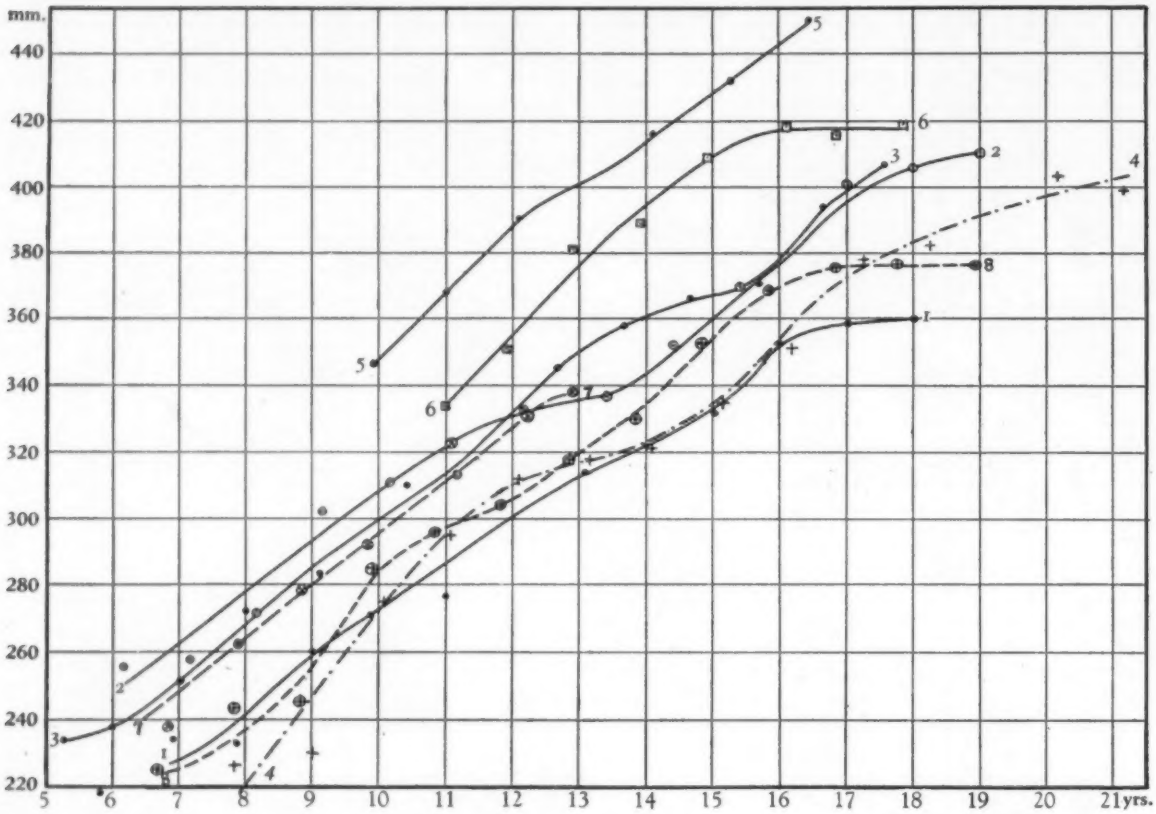


FIG. 24. Individual curves of growth in length of shank (tibia) of 8 boys: 1, J. G., no. 114 (Italian); 2, E. H., no. 1; 3, C. H., no. 83; 4, C. D., no. 30; 5, M. G., no. 95 (acromegalic); 6, R. H., no. 80 (col.); 7, H. T., no. 3 (col.); 8, J. M., I, no. 129.

III. CHANGE WITH AGE OF PERCENTAGE RATIOS AND INDICES OF LINEAR DIMENSIONS

1. GROSS UPPER LIMB RELATIVE TO STATURE

Figure 25 shows the differential growth of the upper limb as related to stature. In early prenatal life (12 weeks) the upper limb is only 34 per cent of stature. It advances in the next three months to 39 per cent and thereafter changes little to birth (data of Scammon and Calkins, 1929: 216). From Kugler's (1932) data the ratio at birth would be about 42.9 per cent for males and 42.1 per cent for females, while Schreiber's (1928) data give 39.4 and 39.8 per cent, respectively. Scammon and Calkins' (1929) data give a ratio of 38.6 per cent. My own data give a mean percentage ratio at birth of 39.8 for males and 40.9 for females, but the rates are based on only about four sets of measurements of fetuses of each sex. I have run the curve through 40.4. At birth the mean upper limb is two-fifths of the total body length (stature).

In my data, soon after birth, as the lower limb elongates rapidly (fig. 19) and the back straightens out (see next section) the mean percentage ratio begins to decrease, falling to about 39.5 at 6 months p. p. in the males and to slightly less in the females. Thereafter for a year and a half the mean ratio increases. An eventual per-

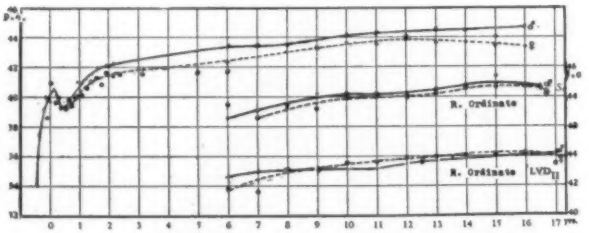


FIG. 25. Standard mass curves of change with age of percentage ratio of length of gross upper limb to stature made up as in fig. 4. Three sets of ordinates. *S*, from data of Saller, Fehmaraner. On the upper curves the observations from the sixth year on and the two prenatal points not on the curves were made by the author.

centage ratio of about 44.6 is attained by the mean males.

Individual babies.—The curves of the change of the ratio from the first into the fourth year depart from the horizontal line that indicates no differentiation. In general there is an increase in the ratio from the first to the fourth year which indicates that the gross upper limb length in babies during this time is growing faster than stature. However, in one case there is no significant increase of the ratio. The ratio into the middle of the first year p. p. stands at about 39 to 41 per cent. In the middle of the fourth year this ratio has increased from 40 to 42.6 per cent. The most rapid change in the slope of the curve occurs during the first and second years. Thus in one curve at 9 months the upward slope with age is one point per year, whereas in the same curve at 3 years the upward slope is only about one-tenth of a point. The curves do not slope uniformly upward with age, but are subject to waves, suggesting that there are periods in which the arms grow in relation to stature now more slowly, now more rapidly.

Individual youth.—Figure 26 gives, for 7 boys, the curves of change with age of the per-

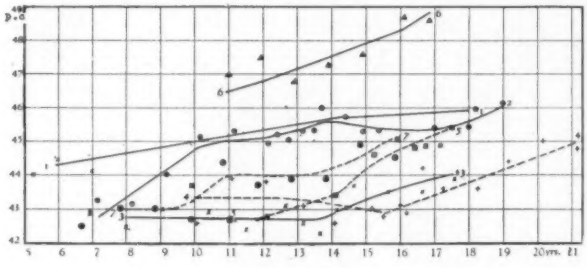


FIG. 26. Individual curves of change with age of percentage ratio of length of gross upper limb to stature for 7 boys: 1, J. G., no. 114 (Italian); 2, E. H., no. 1; 3, C. H., no. 83; 4, C. D., no. 30; 5, M. G., no. 95 (acromegalic); 6, R. H., no. 80 (col.); 7, J. M., I, no. 129.

centage ratio: gross upper limb length divided by stature. In all cases the curves show an increase in the ratio from childhood to puberty. At 7 years the median position of the curves is about 43.5 per cent, which ties up fairly well with the ratio found in babies from the fourth year—between 40 and 43 per cent. By the eighteenth year the median position of the curves has attained 45.5 per cent. This indicates a continued more rapid growth of the gross upper limb than of stature during the juvenile and adolescent periods. This is perhaps to be expected because, although the growth of the upper limb is associated with that of the lower limb, that of the trunk, head, and neck is comprised in the growth of stature and these parts grow slowly. The slope of the curves is sometimes fairly regular, as in nos. 1 and 6. In other cases there are periods of more or less rapid growth. The case of no. 4 is somewhat aberrant in that it slopes downward from 10 to 15 years and then rises to 21 years. The uppermost curve in the series is that of the colored boy. His position is in accordance with expectation, based on the well-known long limbs of the Negro. Here the limbs are growing in length much faster than the trunk, head, and neck. The next lower curve in the series is that of the delicately built Italian boy, in whom also the arms are relatively long. The curve of the acromegalic, no. 5, comes in the lower half of the series, and that of the idiot boy almost in the middle.

Twins and special cases.—Figure 27 gives curves of change with age of the percentage ratio of gross upper limb length to stature for two pairs of twins and some special cases. Curves nos. 1 and 2 are of the monozygotic twin girls. They do not run precisely parallel and are about one-half to one point apart. Curves nos. 5 and 6 of the dizygotic Negro girls lie two and a half points apart. Curves nos. 7 and 9 are of two untreated cretinous children; they seem to lie not far from the median trend of the whole group. No. 8, the treated cretin, lies also not far from the middle of the group. No. 10 is that of a chondrodystrophic girl, and her ratio lies very high, but on the other hand, the ratio of no. 11, a chondrodystrophic man, is extraordinarily low, at about 38 per cent. Evidently in the chondrodystrophics it is the bones of the upper limb of which the growth is most affected in some cases, and the bones of the lower limb in others. The growth curves of the ateliotic girls, nos. 12 and 13, lie near the top of the series, showing that for them the growth of the lower limbs is more interfered with than the growth of the upper limbs. No. 14, that of the microcephalic, lies also near the top of the series, confirming the popular idea of relatively longer arms of the idiots.

Summary.—The upper limb grows much faster than stature before birth; slightly faster after infancy. There is a depression at about 6 to 8 months due to the exceptionally rapid growth of the lower limb (and stature) then.

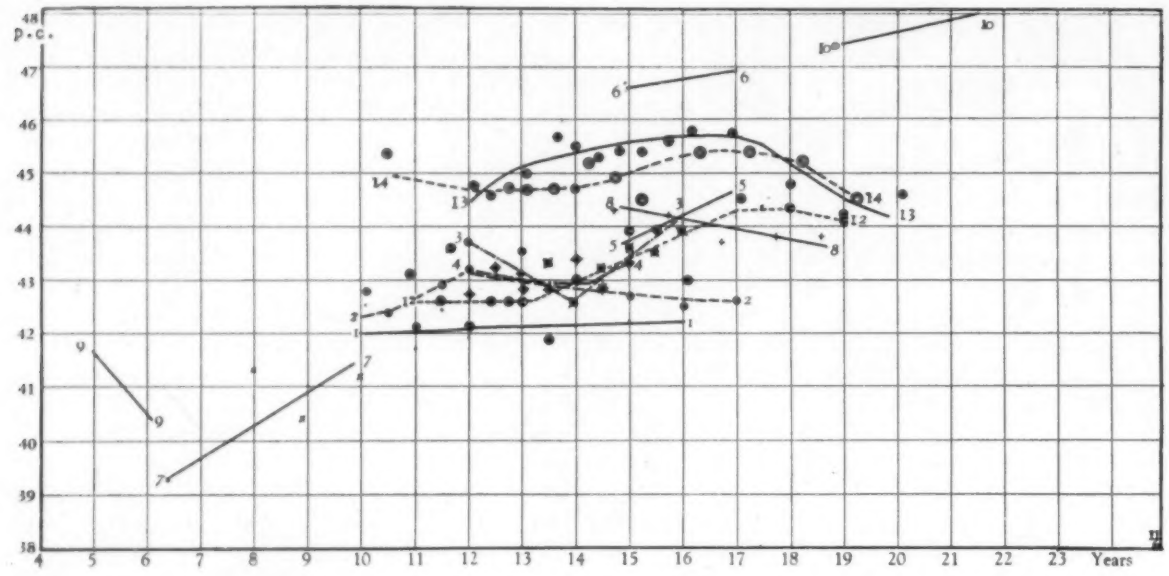


FIG. 27. Individual curves of change with age of percentage ratio of growth of length of gross upper limb to stature in twins and some special cases: 1, 2, A. G. and M. G. (monoz. twins), f.; 3, 4, Ri. O. and Ro. O. (monoz. twins), m.; 5, 6, E. D. and S. D. (dizygotic twins), f.; 7, A. A., cretin, f.; 8, C. B., treated cretin, f.; 9, H. L., cretin, m.; 10, E. S., chondro., f.; 11, W. S., chondro., m.; 12, M. C., atel., f.; 13, A. W., micros., f.; 14, V. R., microcephalic, m.

2. GROSS UPPER LIMB RELATIVE TO SITTING HEIGHT

The importance of this ratio is that it reveals the age changes relative to a part of the body where growth is not closely correlated with the arm length—is an independent variable. This independence renders the ratio more significant than a comparison of gross upper limb length with total stature, which involves leg length. We have to do with a relative growth of upper limb in comparison with trunk somewhat in the sense of Huxley's relative growth except that the dimensions are linear in this case.

Sitting height is self-explanatory. This dimension corresponds closely to the length of the body of any mammal, from the head between the ears to the buttocks. Unfortunately, we have few comparative data on the way this ratio differentiates with age either in man or any other primate. In the new-born chimpanzee the length of the upper limb is found by Schultz (1940: 23) to be 170 per cent of trunk height, to increase in the mean infant to 179, and to diminish to 172 in the adult. Since trunk length is of the order of 60 per cent of sitting height, 170 per cent of trunk height would correspond to about 100 per cent of sitting height.

Figure 28 gives the percentage ratio of gross upper limb length to sitting height in masses of children. A few points in fetal development computed from means given by Scammon and Calkins indicate that this ratio increases rapidly from about 45 per cent at 3 months of gestation to about 58 per cent at birth, just as it does in the chimpanzee, but it reaches less than 60 per cent of the ratio of that anthropoid.

After birth the ratio may increase a bit in the male but slows down in the middle of the first year p. p. A study of sitting height in babies during the first year shows that an increase of gradient frequently occurs at about 5 to 7 months. This is about the time when a baby first sits in a chair, "holding head erect and steady" and capable of turning the head freely (Gesell and Thompson, 1938: 166). Evidently

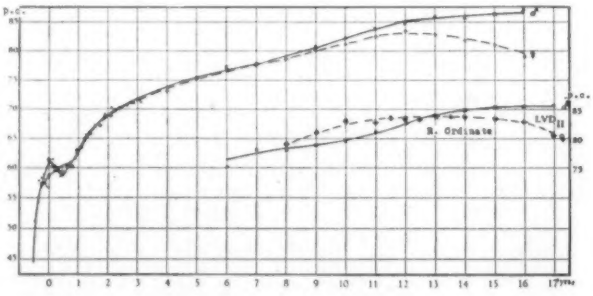


FIG. 28. Standard mass curves of change with age of percentage ratio of length of gross upper limb to sitting height. Two sets of ordinates.

this is a period when the muscles of the back and neck are developing rapidly.

Sexual.—In general, the difference between boys and girls in the ratio of upper limb to sitting height is slight, until after puberty, when the upper limb of the girl may become absolutely shorter. (See fig. 1.)

Social.—The percentage ratio in idiots is strikingly higher than in the control, especially in pre-adolescents. One low-grade boy has the ratio of 94 per cent. After adolescence the upper limb tends to grow slowly, so that it falls behind in its relation to sitting height.

Racial.—The percentage ratio of gross upper limb length to sitting height is greatest in the colored children, least in the Nordics; the curve of Mediterraneans lies between.

Summary.—The percentage ratio of the length of the gross upper limb to sitting height increases rapidly till the middle of the first year p. p.; slowly for 6 months; then more rapidly until a general reduction of slope occurs in late childhood. The sexes are alike in this differentiation process. In babies the differentiation proceeds fairly uniformly, with occasional cessation. During post-adolescence the differentiation curve advances at varying rates, and its smooth progress is occasionally interrupted. Rarely, except for the idiot group, is the slope downward as found in chimpanzees. In twins the differential curves lie parallel and close together. In cases of abnormal long-bone development this ratio may be quite aberrant. Fraternities differ in homogeneity of the curves.

3. GROSS UPPER LIMB RELATIVE TO GROSS LOWER LIMB (INTERMEMBRAL INDEX)

The limbs are homologues, yet in the course of evolution their relative length has become very different in different species, adapted as they are to different conditions of life. The anterior and the posterior appendages are frequently adapted to the performance of very different functions, as in rabbits, moles, kangaroos, gazelles.

The percentage ratio of length of upper to lower limb is called the intermembral index. I have referred to age changes in this index in earlier papers (1934; 1935).

Individual babies.—Figure 29 shows the curves of changing index with age. They start at birth to 2 months at 90 to 96; at one year p. p. they range from 84 to 95. Thus in infants the lower limbs are at first hardly longer than the upper limbs. A sudden rise in the index occurs between 13 and 18 months, in connection with erect posture. By 3 years the index has fallen to points between 90 and 85.

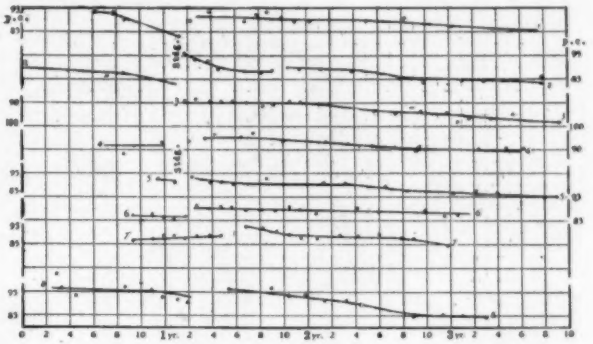


FIG. 29. Individual curves of change with age of percentage ratio of gross upper limb length to gross lower limb length for 8 babies as follows: 1, E. K., f. no. 1; 2, B. B., f. no. 2; 3, C. C., f. no. 3; 4, A. L., f. no. 5; 5, R. N., f. no. 6; 6, F. M., f. no. 7; 7, M. M., f. no. 8 (twin sister to no. 7); 8, T. P., m. no. 9. 7 ordinate scales. Breaks at 1 to 1:5 years due to change of posture in measurement.

Individual youth.—Figure 30 shows for individual boys how devious are the changes of the intermembral index with age. In general the curves slope downward until 13 or 15 years and then trend upward with equally marked slope of sometimes a point per year. This is the U-shaped differentiation curve.

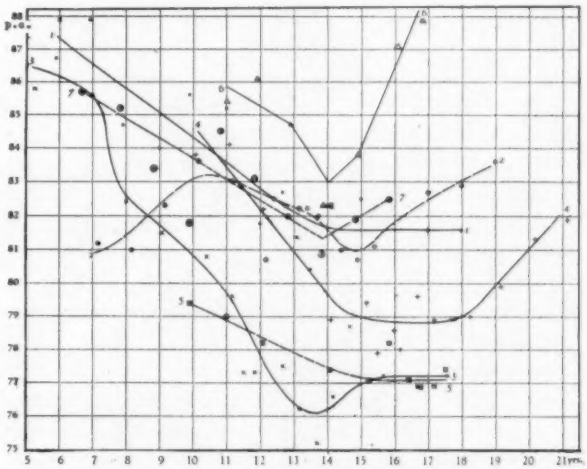


FIG. 30. Individual curves of change with age of percentage ratio of gross upper limb length to gross lower limb length for 7 boys: 1, J. G., no. 114 (Italian); 2, E. H., no. 1; 3, C. H., no. 83; 4, C. D., no. 30; 5, M. G., no. 95 (acrom.); 6, R. H., no. 80 (col.); 7, J. M., I, no. 129.

Summary.—The curve of intermembral index is irregular for 6 months before and 18 months after birth, owing to such things as release from intra-uterine restraint and shortening of lower limbs by gravity on standing. Thereafter the index decreases. Girls have a smaller index than boys, and Mongoloid dwarfs a smaller one than control. Negroes have a high index. An acromegalic has a low index; chondrodystrophics have high ones. Different fraternities have different types of change of slope of the index curve with age.

4. NET UPPER LIMB RELATIVE TO STATURE

As this ratio advances with age, it is constantly between 0.25 and 0.5 point greater in males than in females; that is, the net upper limb length is not only absolutely (fig. 4) but relatively less in girls than in boys. For boys the mean percentage ratio increases from 31.8 at 6 years to 33.2 at 16 years.

Social.—The relative net upper limb length is much less in Mongoloid dwarfs than in the control, and this difference increases with age; in fact, percentage mean relative limb length decreases from about 32 at 7 years to 30.5 at 19 years and keeps about 1.3 to 2 points below standard.

Racial.—Of the three groups considered in this study, the Negro has by far the longest relative net upper limb. The means, in my data, run from about 32.8 at 9½ years to 33.7 at 14½ years; while the corresponding Nordic group stands about 0.5 point below, and the Mediterranean group is intermediate but close to the Nordic. According to Weninger (1927: 113), Negroes of French West Africa have a percentage mean relative net upper limb length of 34.8.

Summary.—Since the net upper limb grows much faster than stature, the curve ascends steadily after 6 months. The sexes are alike in this respect to 12 years. Especially in girls the curve tends to slope downward after adolescence.

5. NET UPPER LIMB RELATIVE TO NET LOWER LIMB (NET INTERMEMBRAL INDEX)

The net intermembral index is in many ways better than the gross index for comparisons because of the special plantigrade adaptation of the human foot. This index enables a comparison to be made between the two long bones of the upper and of the lower limbs, both in the individual and between races and species. As Martin (1928: 427) says:

> A comparison of the upper and lower extremities (without the distal segments) reveals the extraordinary development in length of the lower extremity of man in contrast with all other Primates, conditioned by his vertical posture.

General.—The curve of change of the intermembral index with age is a complicated one (fig. 31). The prenatal section, based on data of Scammon and Calkins (1929), does not tie up well with the postnatal portion. They find an

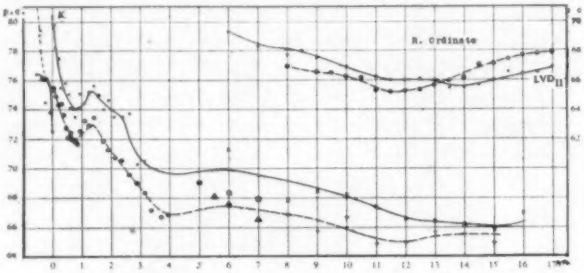


FIG. 31. Standard mass curves of change with age of percentage ratio of net upper limb length to net lower limb length. Scale of ordinates to LVD curves on the right. K, Kugler, neonates. Broken line: alternative curve based on a few original observations.

index of 72.5 at term. In my data the male neonate has an index of 79.8. Kugler's data give an index of 80.4 for males.[3] During the first 10 months p. p. the index declines to 74, owing to the exceptionally rapid growth of the lower limb. At the transition to standing, the lower limb is shortened by gravity and the curve ascends on the average for a few months. After this episode the curve descends to maturity, since the lower limbs grow faster than the upper limbs. But the rate of descent is variable. In the third year it is precipitous. This is a period when the lower limb is increasing rapidly in length (fig. 16). In the sixth year the mean ratio undergoes practically no change. This is a period of low growth gradient in both limbs. Then may follow a slight downward slope. On the whole, childhood and juvenility are periods of especially rapid growth of the lower limbs (fig. 1) or slow growth of the upper limbs (fig. 4), or both. The result is an index of 66, or less, at maturity.

[3] The gross intermembral index of 95 or more in neonates (sometimes cited) is due to the inclusion in the ratio of the long hand and not of the metatarsals.

Comparative.—Light on the significance of the human curve of differentiation of the intermembral index during development may be gained by considering that index in other Primates and, indeed, in mammals generally. Tables 2 and 4 give the results of this survey.

Adult man has a smaller net intermembral index than any of the anthropoids. The neonate is most like an anthropoid in this respect. Also the new-born chimpanzee has a higher index than the adult, but the lowering of the index in the adult chimpanzee is small, 7 points (Schultz, 1940: 23), while in *Homo* the reduction is more nearly of the order of 20 points. The difference between the intermembral indices of chimpanzee and human is due to the slow growth in length of the upper limb and the rapid growth of the lower limb in *Homo*. In the chimpanzee both relative lengths change little during development (table 1).

TABLE 1

CHANGE IN PERCENTAGE RATIO OF RELATIVE LENGTH (TO STATURE) OF GROSS UPPER AND GROSS LOWER LIMBS IN PASSING FROM 4 YEARS TO MATURITY IN CHIMPANZEE AND HUMAN

Change in:	Chimpanzee*	*Homo*
Rel. lgth. of gross upper limb	−9	+18 (fig. 25)
Rel. lgth. of gross lower limb	−13	+55

* Schultz (1940: 23).

The difference in intermembral index between anthropoid and *Homo* is thus due more to the luxurious growth of the lower limb of *Homo* than it is to the restricted growth of the upper limb.

Is the low intermembral index of the kangaroo due to an abnormally large growth of the posterior appendages or to a reduction of the anterior appendages, or both?

To compare more in detail the change with age of the intermembral index of the four anthropoid apes, I have computed means of the index for juvenile, medium adults, and large adults from data furnished by Schultz (1930: 430–435) (table 2).

This survey of the anthropoid apes shows (1) the great arm length of the orang-utan and gibbon; (2) the restriction of the arm length in the gorilla and especially the chimpanzee; (3) the possible reduction of the intermembral index at the time of the adolescent spurt (chimpanzee and gorilla) as in man.

Again, there is some evidence that in the arboreal adolescent group (orang-utan, medium)

TABLE 2

MEANS OF INTERMEMBRAL INDEX FOR FOUR ANTHROPOID SPECIES OF THREE AGES: SMALL (S), MEDIUM (M) AND LARGE (L)

Species	Frequency	Small (young)	Medium	Large
Chimpanzee	(s, 7; m, 7; l, 10)	110	105	109
Gorilla	(s, 8; m, 20; l, 10)	117	115	117
Orang-utan	(s, 4; m, 10; l, 12)	143	144	144
Gibbon	(s, 7; l, 10)	122		128

the humerus grows faster than the femur (table 3). Of course, the high index in orang and gibbon is associated with their arboreal life.

As the anthropoids became terrestrial, the upper limb length shortened as compared with the lower limb. Thus, as Mollison (1910: 111) shows the net upper limb is 189 per cent of the anterior length of trunk in gibbon; 160 per cent in orang; 134 per cent in gorilla; 123 per cent in

TABLE 3

PERCENTAGE RATIOS OF HUMERUS LENGTH TO FEMUR LENGTH AT THREE AGES *

	Young	Medium	Large
Chimpanzee	107	101	106
Gorilla	115	114	119
Orang-utan	127	132	126
Gibbon		116	116

* Based on data from Schultz (1930).

chimpanzee; and 116 per cent in the Badeners of Germany. The net lower limb does not show this shortening of the upper limb. Thus the net lower limb is 131 per cent of the anterior trunk length in the gibbon; 113 per cent in chimpanzee; 113 per cent in gorilla; 111 per cent in orang; and 159 per cent in man. The length of the lower limb has *increased* relatively in man.

Table 4 shows how the intermembral index varies from the high index (200 or more) found in bats to the low index (33) found in jumping mice and kangaroos. Since man is neither a flying animal nor, like the gibbon, a denizen of the treetops, he has low intermembral index of 66, like certain of the jumpers. The region of the scale between 80 and 100 is the most crowded of all. It may be thought of as the most typically mammalian.

Discussion.—The net intermembral percentage ratio of man at birth is approximately 80, 30 less than in the chimpanzee. It descends during

TABLE 4

NET INTERMEMBRAL INDEX OF VARIOUS MAMMALIA (PERCENTAGE RATIO OF LENGTH OF UPPER TO LOWER LIMB)

Promops nasutus (bat)	219
Eptesicus fuscus (big brown bat)	206
Pteropus giganteus (Malay bat)	199
Pteropus vamp (bat)	190
Bradypus (sloth)	160
Zalophus californianus (sea lion)	115
Choloepus hoffmanni (sloth)	110
Scalopus aquaticus (mole)	109
Giraffa camelopardalis (giraffe)	108
Odobenus rosmarus (walrus)	102
Dasypus peba (armadillo)	101
Camelus bactrianus (Bactrian camel)	97
Elephas maximus (Asiatic elephant)	97
Erethizon epixanthum (yellow-haired porcupine)	97
Ursus horribilis (grizzly bear)	92
Felis tigris (tiger)	90
Macacus rhesus (rhesus monkey)	90
Myrmecophaga (anteater)	90
Vulpes vulpes (fox)	88
Cynopithecus niger (baboon)	88
Phascolomys mitchelli (wombat)	86
Petaurus (flying phalanger)	86
Didelphis virginiana (oppossum)	84
Cebus capucinus (monkey)	82
Dicotyles torquatus (collared peccary)	81
Cervus canadensis (wapiti)	80
Tapirus americanus (tapir)	79
Geomys bursarius (gopher)	77
Dasypus (armadillo)	76
Manis (pangolin)	74
Hydrochoerus (capybara)	73.5
Sylvilagus (hare)	71
Castor fiber (European beaver)	69
Homo africanus (man)	67
Homo europeus (man)	66
Zapus hudsonius (jumping mouse)	55
Dipodomys (kangaroo rat)	52
Tarsius (tarsier)	46
Onychogale unguifer (nail-tailed wallaby)	43
Macropus robustus (kangaroo)	42
Macropus rufus (red kangaroo)	42
Pedetes cafer (jumping hare)	39
Scirtopoda orientalis (jumping mouse)	33

development, eventually reaching 66, while in the chimpanzee the reduction is relatively slight. Thus the developmental changes in the two species are in the same direction, only they go farther in *Homo*. As appears from an earlier discussion (table 1) the decrease in the intermembral index is due rather to a more rapid growth of the lower limb than to a slowing up of growth of the upper limb. Man has gained his limb proportions by an exceptionally fast growth of the lower limbs compared with a more normal growth of the upper limbs.

Summary.—The net intermembral index decreases with age after 1½ years, since the lower limbs grow faster than the upper. Typically the index is smaller for girls than for boys, at least to puberty. The idiot children have a large index (long upper limb), and the Mongoloid dwarfs have a small index. The Negroes have a large index. Boys show a minimum just before the adolescent spurt. Chondrodystrophics have a very high index. The luxurious growth of the lower limb in *Homo* vs. *Pan* is responsible for the fact that from 4 years to maturity the index is low in man and high in the chimpanzee.

6. "ARM" LENGTH RELATIVE TO GROSS UPPER LIMB LENGTH

This is the percentage ratio of the "arm" of anatomists (or humerus) to total gross upper limb length. The curve of age-change shows how this segment of the arm differentiates.

Figure 32 shows that during gestation the "arm" grows in length more slowly than the rest of the limb; the percentage ratio falls from 43 to

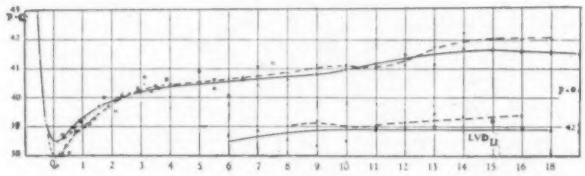


FIG. 32. Standard mass curves of change with age of percentage ratio of "arm" length to gross upper limb length.

38.5 at birth. Immediately after birth the mean percentage ratio increases rapidly during 2 years to 40, and despite continued growth of the arm as a whole remains thereafter within 2 points of that ratio.

The mass curve of change with age of the ratio of humerus length to total upper limb length is remarkably constant after 3 years. But this is owing largely to the fact that changes in proportion occur in different individuals at different ages, so that the ratio is flattened out. However, there is a tendency for the humerus to grow especially fast at the time of the adolescent spurt when that occurs between 12 and 16 years. The bone-forming processes of cell-proliferation and collagen formation are accelerated by hormones at this time and more in Negroes than in Whites.

7. "ARM" LENGTH RELATIVE TO NET UPPER LIMB LENGTH

This ratio is of interest as indicating the relative growth of the two chief bones of the upper limb—humerus and radius. Except that the humerus lies nearer the trunk, these two bones would seem to be equally supplied by nutrition and stimulating hormones. Do they grow at similar speed?

The humerus constitutes at birth the smallest linear fraction of net arm length that it does at any age—53 per cent. Typically this percentage ratio rises in 2 years to 55.5, and at adolescence it reaches 56.2 in the male, 56.5 in the female. It sometimes increases after 16 years. In different fraternities it lies between 54.5 and 58.5.

8. "ARM" LENGTH RELATIVE TO THIGH LENGTH (FEMORO-HUMERAL INDEX)

We have considered the intermembral index between limbs as a whole. It is worth while now to consider the comparative growth between homologous segments of the upper and lower limbs; especially of their bones—in this instance humerus and femur.

General.—The mean differentiation curve (fig. 33) between humerus and femur shows little of interest except in the circumnatal period. The neonatal index, as found by different observers, ranges from 66 (Scammon and Calkins, 1929, fetuses) to 84 (Kugler, 1932, males). My own finding is 78.7 for males and 72.5 for females. This variation is doubtless due to differing condi-

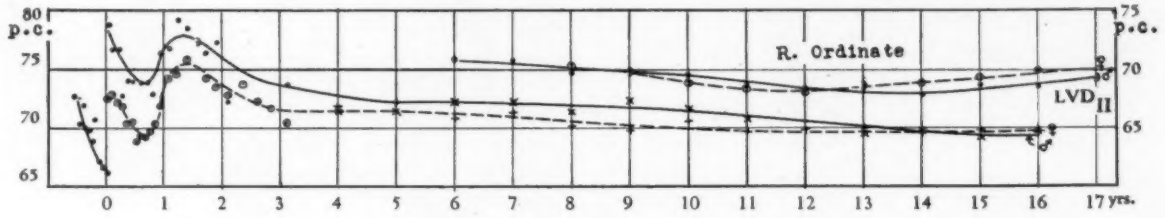


FIG. 33. Standard mass curves of change with age of percentage ratio of femoro-humeral index. Ordinates at the right for D2 series.

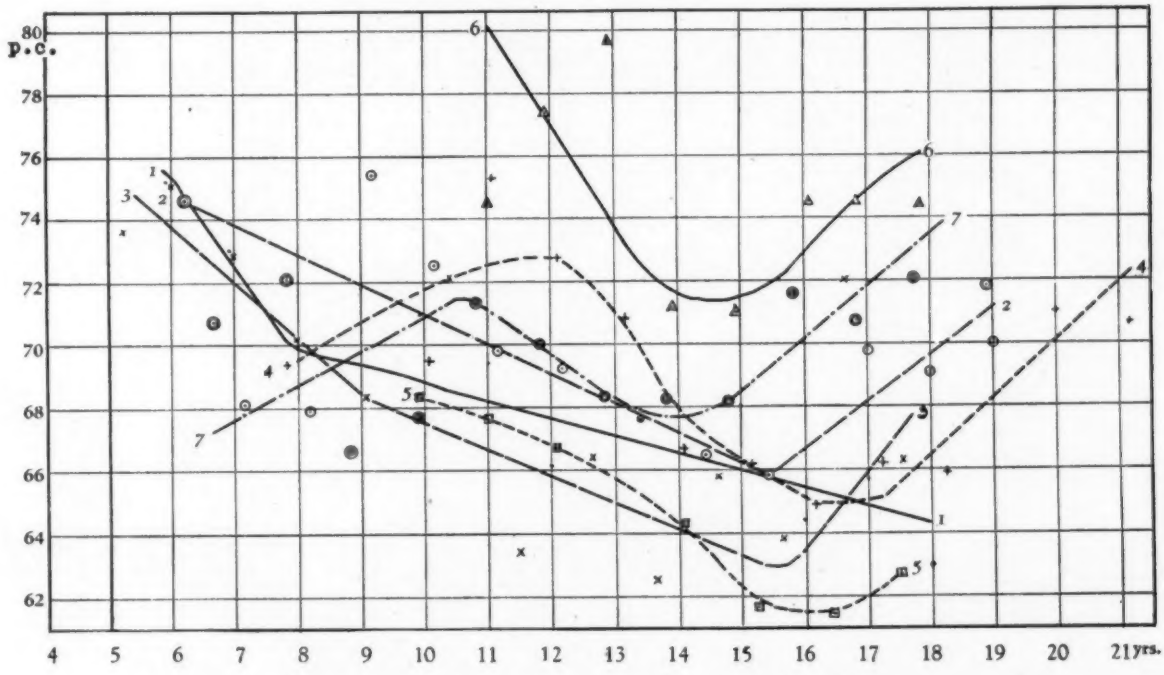


FIG. 34. Individual curves of change with age of percentage ratio of femoro-humeral index for same 7 boys as in fig. 30.

tions of the subjects and differing landmarks. During the first 6 months p. p. there is a downward slope indicating the especially rapid growth of the femur. The up-slope to 16 months is partly the reaction from the preceding minimum (slowing down of growth of femur), partly the shortening of the femur on standing. From this maximum the slope is downward about 0.6 point per year to an index of 64 at 16 years.

Racial.—The Nordics and Mediterraneans have a similar femoro-humeral index; but the Negro children have a larger index (proportionally long humerus)—about 72 in place of 70 of whites.

The femoro-humeral index as computed from bones gives the highest index for Italians (72.9), next for Fiot Negroes and the Salado Indians, 72.6. The smallest index is given by the Andaman islanders (Flower, 1880), 68.6.

Individual youth (fig. 34).—All curves show a downward slope between 12 and 14 years and a rising slope after 16 years (excepting no. 1). This can mean only one thing: in connection with the adolescent spurt the femur first grows faster and later the humerus. The two bones do not accelerate at the same rate at the same time. The curve of the acromegalic (no. 5) lies at the bottom, because of his long femur, while that of the colored boy (no. 6) is at the top—he has a long humerus. Curve no. 1 is remarkable as showing no up-turn to 18 years.

TABLE 5

MEAN FEMORO-HUMERAL INDEX (PERCENTAGE RATIO OF LENGTH OF HUMURUS TO THAT OF FEMUR) IN VARIOUS ADULT MAMMALS

Pteropus (bat)	180
Zalophus (sea lion)	167
Bradypus (sloth)	161
Desmodus (bat)	160
Nycteris (bat)	159
Eptesicus (bat)	158
Promops (bat)	156
Odobenus (walrus)	151
Noctilio (bat)	134
Hylobates conc. (gibbon)	132
Pongo pygnaeus (orang)	130
Scalops (mole)	122
Symphalangus	121
Gorilla (gorilla)	119
Phoca (seal)	116
Hylobates hool. (gibbon)	115
Macacus nemestrinus (macaque)	115
Potamogale (aquatic insectivore)	113
Pan (chimpanzee)	105
Solenodon (insectivore)	104
Erinaceus (hedgehog)	103

TABLE 5—*Continued*

Choloepus (sloth)	102
Dasypodidae (armadillos)	98
Hemicentetes (insectivore)	96
Tayassu (peccary)	95
Sus (pig)	95
Canis (dog)	94
Vulpes (fox)	93
Rhinoceros (rhinoceros)	93
Giraffa (giraffe)	93
Papio (baboon)	92
Felis (cat)	91
Macacus (macaque)	91
Conepatus (skunk)	90
Nycticebus (lemur)	89
Nasalis larv. (proboscis monkey)	89
Cynopithecus (monkey)	89
Macacus fuscus (macaque)	88
Acanthion jav. (porcupine)	88
Chimarrogale (insectivore)	88
Didelphis (opossum)	88
Erethizon (porcupine)	87
Ursus americanus (black bear)	87
Elephas (elephant)	85
Manis (pangolin)	85
Hydrochoerus (capybara)	84
Marmota (woodchuck)	84
Haplodon (mountain beaver)	84
Camelus (camel)	84
Dasyprocta (agouti)	84
Cervidae (deer)	83
Petaurus (flying phalanger)	82
Cercocebus mona (monkey)	82
Dendrolagus (tree kangaroo)	82
Cebus (monkey)	81
Colobus (monkey)	80
Eliomys (dormouse)	80
Pithecus (monkey)	80
Cercopithecus cer. (monkey)	80
Tapirus (tapir)	79
Sylvilagus (hare)	79
Bubalus (buffalo)	79
Lepus (hare)	79
Phascolomys (wombat)	79
Geomys (gopher)	78
Castor (beaver)	78
Semnopithecus (monkey)	78
Cercocebus sp. (monkey)	78
Myrmecophaga (anteater)	75
Pygathrix (monkey)	75
Equus (horse)	74
Microtus (field mouse)	73
Lemur (lemur)	70
Cercopithicus petaur. (monkey)	69
Thryonomys ("ground rat")	66
Zapus (jumping mouse)	66
Homo europaeus (man)	64
Macropus giganteus (kangaroo)	62
Napaeozapus (jumping mouse)	59
Dipodomys (kangaroo rat)	55
Macropus robustus (kangaroo)	52
Onychogale (nail-tailed wallaby)	49
Dipodomys spin. (kangaroo rat)	45
Pedetes (jumping hare)	44
Alactaga (jerbod)	41
Scirtopoda (jumping mouse)	37

Comparatives.—In all other Primates the femoro-humeral index is higher than in *Homo*. In adult monkeys the indices vary around 85. In the adult apes the index is generally over 100—chimpanzee 105, gorilla 119, gibbon 132, orang-utan 130. Thus in the human neonate we have an index (78) approaching that found in apes; eventually it will fall to 64 or less.

The list of indices of various mammals (table 5) shows the position of *Homo* in the series. From the table it appears that, in the percentage relation of humerus to femur, i. e., the index, *Homo* at 64 stands rather low in the series—among the jumping mice and kangaroos at 66 and 37. The monkeys lowest in the series are the macaques, 88, and certain other long-tailed monkeys (*Cercopithecus*, 80 and 69). The apes are rather distant, 132, 130 and 105. The bats with their extraordinarily long humeri lie at the upper extreme (180 to 134). With his reduced "arm" and highly developed thigh, man has left the brachiating group to become a runner on his hind limbs and a leaper.

Summary.—Except for a striking series of readjustments during infancy, the percentage ratio of humerus to femur undergoes no change after 3 years of age except a reduction from 73.5 to 64.5. The femur gradually becomes the dominating long bone in the body. This development places *Homo* between the running and the jumping animals, far removed from the brachiating apes. Man's fetal index is high, as is that of the apes; but whereas the apes retain to adulthood a high index of 100 to 130, man's index falls so as to place him rather low in the mammalian series, among the terrestrial forms.

9. FOREARM (RADIUS) RELATIVE TO "ARM" (BRACHIAL INDEX)

The brachial index is the percentage ratio of the length of the distal to that of the proximal upper limb segment—the radius to the humerus. Were there no differential growth, the index would remain constant during development. That it does not remain so is evidence of differentiation.

Figure 35 shows the mean changes with age of the brachial index. The prenatal part of the curve is based on the work of Scammon and Calkins and also on that of Schultz. The index rises from 75 at 12 weeks, to 82 at 38 to 40 weeks.[4] One may compute from their means of lengths of "arm" and of forearm the index at birth. It is 87.5. From living neonates I get an index of 90.6 for boys and 91.9 for girls. Kugler (1932: 479) found indices of 93.2 and 94.5, respectively. Slight differences of technique would explain the differences. Roughly, at birth, the mean child of Nordic stock has a radius of length about 90 per cent of that of the humerus.

Racial.—While between the Nordics and the Mediterraneans no clear difference in racial index appears, the Negro children had a definitely higher index (about 5 points higher)—at 15 years 84, as contrasted with 77 for Whites of the same group. This difference agrees with that found in comparing adult Europeans and Negroes of Africa. Thus, while for adult Europeans indices of 75 to 80 have been found, for adult Negroes indices of 78 to 93 are cited by Martin (1928: 395). The radius in the Negro exceeds that in the Whites more than the

[4] Two fetuses, aged 16 weeks, measured by me gave an average index of 93.

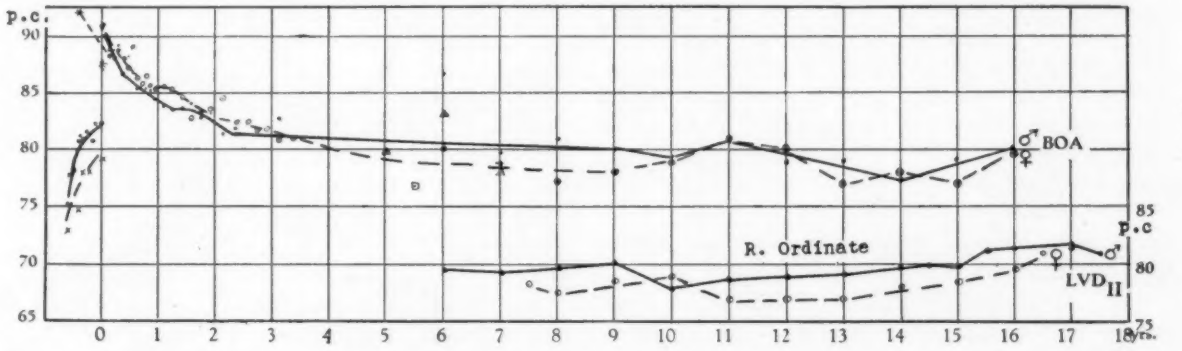


FIG. 35. Standard mass curves of change with age of percentage ratio of forearm length to "arm" length (radius/humerus). Scale of ordinates at right refer to LVD 2 curves. The asterisks (*) at 4 months and at zero abscissa are based on fetuses measured at Department of Embryology by the author.

humerus does. The Negro has a long upper limb and especially a very long distal segment of that limb. This is a brachiating trait—marked in the orang and gibbon. It is also an early ontogenetic trait. As long ago as 1862, Broca called attention to the long radius of Negroes.

Individual babies.—As figure 36 shows, the brachial index of infants progresses with age tumultuously. There is the steep downward

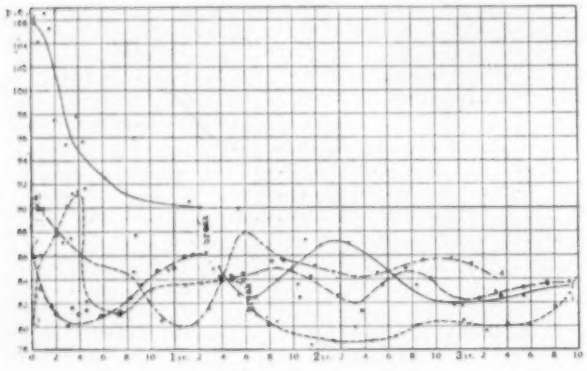


FIG. 36. Individual curves of change with age of percentage ratio of forearm length to "arm" length (radius/humerus) for 4 babies: 1, E. K., no. 1, f.; 2, C. C., no. 3, f.; 3, R. N., no. 6, f.; 4, T. P., no. 9, m. Breaks at 1:2 yrs. due to first assumption of standing posture.

slope of three of the four curves from birth to 4 months and in the case of 2 to 14 months. In all cases the age-change curve is of the wavy type, suggesting alternating periods of more and of less rapid growth of the radius relative to that of the humerus.

Individual youth.—When the curves of age-change are plotted on a large scale for boys beginning at ages of 6 to 8 years and ending at 18 to 21 years (fig. 37), wave-like irregularities appear which can hardly be due to technical difficulties and are most probably to be ascribed to variations (even alternations) of growth in the radius and humerus. A maximum is found at the ages of 15 years (once); 16 (twice); 17 (twice); and 19 (once). These are ages of the adolescent spurt.

Comparative.—To understand the human brachial index it is necessary to consider the differences of the index throughout the mammalian series. To determine this index I measured the radius and humerus of adults of the species listed in table 6. This I was able to do with the cordial co-operation of two curators of mammals: Mr. Gerrit S. Miller of the U. S. National

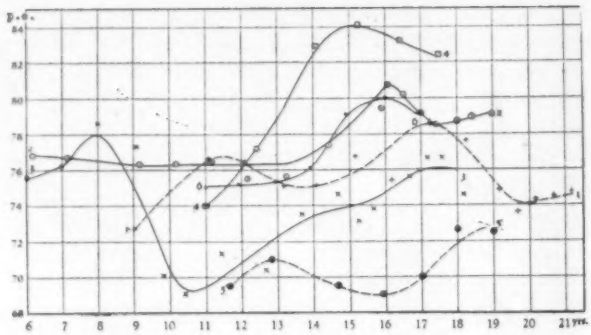


FIG. 37. Individual curves of change with age of percentage ratio of forearm length to "arm" length (radius/humerus) for 6 boys: 1, E. H., no. 1; 2, C. D., no. 30; 3, C. H., no. 83; 4, M. G., no. 95 (acromegalic); 5, P. M., I, no. 10; 6, R. H., no. 80 (col.).

Museum and Dr. H. E. Anthony of the American Museum of Natural History, whose assistance is cordially acknowledged.

From this table the following facts emerge: The longest forearm segment as compared with "arm" is found in bats (indices from 205 to 138). The relatively shortest forearm is found in certain burrowing or swimming forms: *Potamogale*, an insectivorous "otter," 58; *Manis*, the scaly anteater, 66; the peccary, the pig, and *Capybara*, a giant guinea pig.

The significance of this distribution is not abstruse. Lone distal segments give a great reach in the more mobile portion of the appendage. Such a reach is especially useful in the wing. It has been shown elsewhere that a long distal segment in the hinder appendage is especially useful in making great leaps and is accordingly found in jumping species. On the other hand, in animals that burrow and swim the proximal segment is the longer. Thus, *Potamogale* has the swimming habits of an otter, *Manis* makes tunnels, the peccary and *Capybara* are excellent swimmers and the pig and allies root the ground with the snout, pushing with their powerful anterior appendages. Again, the walrus (index 76) is a powerful swimmer. In burrowers the main force is exerted by the powerful proximal segment of the anterior limb. In swimmers the appendage is used as a whole, and the proximal portion is that to which the force is especially applied.

The long distal segment of the fore-limb in jumpers is something of a surprise. Of course the long distal segment of their hind-limbs is a physiological necessity. It is possible that the resemblance between the two indices in jumpers

TABLE 6

THE BRACHIAL INDEX (PERCENTAGE RATIO OF LENGTH OF RADIUS TO THAT OF HUMURUS) OF ADULT SPECIMENS OF VARIOUS MAMMALIA, MOSTLY ONE OR TWO INDIVIDUALS OF EACH

Noctilio (bat)	205
Desmodus (bat)	158
Giraffe	155
Nycteris (bat)	149
Pteropus vamp. (bat)	138
Dipodomys (jumping mouse)	130
Napæozapus (jumping mouse)	130
Macropus (kangaroo)	130
Equus grevi (horse)	115
Camelus (camel)	115
Gibbon	114
Myrmecophagus (anteater)	113
Scalops (mole)	113
Choloepus (sloth)	113
Lepus (hare)	111
Pygathrix (monkey)	108
Alactaga (Siberian jumping rat)	107
Papio (baboon)	106
Zapus (jumping mouse)	106
Dasyurus (terrestrial marsupial)	104
Cervus (deer)	104
Eliomys (dormouse)	103
Dendrolagus (tree kangaroo)	102
Cercopithecus (monkey)	102
Microtus (vole)	101
Cynopithecus (baboon)	101
Nycticebus (lemur)	101
Semnopithecus (monkey)	100
Orang	100
Erethizon (porcupine)	99
Pithecus (monkey)	98
Nasalis (monkey)	98
Lemur (lemur)	97
Cercocebus (monkey)	97
Cebus (monkey)	97
Chimarrogale (vole)	97
Bubulus caffer (ungulate)	96
Didelphus (opossum)	96
Castor (beaver)	95
Canis fam. (dog)	94
Phascolomys (wombat)	93
Chimpanzee	93
Colobus (monkey)	93
Bradypus (sloth)	92
Vulpes (fox)	91
Pedetes (jumping hare)	91
Geomys (gopher)	90
Felis leo (lion)	89
Ursus Amer. (bear)	88
Haplodon (sewellel)	88
Phoca vit. (seal)	87
Hemicentetes (insectivore)	86
Sylvilagus (hare)	86
Conepatus (southern skunk)	86
Erinaceus (hedgehog)	84
Tapirus (tapir)	84
Marmota (woodchuck)	84
Rhinoceros	84
Solenodon (insectivore)	83
Elephas (elephant)	82

TABLE 6—*Continued*

Homo (man)	80
Gorilla	80
Thryonomys (ground pig)	80
Odobaenus (walrus)	76
Dasypus (armadillo)	72
Acanthion java (old world porcupine)	71
Capybara (giant guinea pig)	70
Sus scrofa (hog)	70
Tayassu (peccary)	68
Manis (scaly anteater)	66
Potamogale (insectivorous otter)	58

of front and hind-limbs is related to the correlation in variation in length of the two sets of appendages which is known to be high. Between the length of man's femur and humerus the correlation is 0.85.

In Old World monkeys the brachial index is intermediate (108 to 97), just as is the crural index (102 to 97). In the gibbon, which uses its anterior appendages with extraordinary celerity, the brachial index is high (114) while in the more nearly terrestrial gorilla the index is, as in man, low (80). The orang (100) and chimpanzee (93) are intermediate in brachial index.

In fast runners the crural index is low (*Equus*, 73) while the brachial index is relatively high (115). At the first sight this is somewhat unexpected. But it must be kept in mind that it is the hind leg which plays the major part in determining the running gait and pushes the animal forward. The giraffe has an extraordinarily large brachial index (155) connected in some way with its long anterior limbs.

The three-toed sloth, *Bradypus*, which has a remarkably large crural index (174), has only a mediocre brachial index (92). However, the two-toed sloth, *Choloepus*, has a fairly large brachial index (113), placing it near the gibbon. Perhaps *Choloepus* is the more perfectly adapted to the arboreal life.

It appears that those lower mammals that have a semi-erect posture like the kangaroo and the jerboas and *Alactaga* have a higher brachial index than the terrestrial quadrupeds from which they evolved. Compare *Macropus*, 130, with the wombat, 93; and *Alactaga*, 107, with *Eliomys* (dormouse) 103. *Homo*, on the other hand, has a brachial index (80) which is not greater but less than that of the lemur or *Cebus* (97) which are more nearly quadrupedal. However, *Homo* passes, in ontogeny, through a stage with relatively large brachial index (90 at birth). The large brachial indices of the kangaroo and

Alactaga are perhaps correlated with their large crural indices as man's low brachial index (80) is correlated with his low crural index. Were man a jumping animal, instead of a running animal, he would probably have had a longer forearm (fig. 38).

The human infant like that of anthropoids has a large brachial index (90) at birth. It becomes reduced in the adult to 80. The chimpanzee has a brachial index of about 88 at

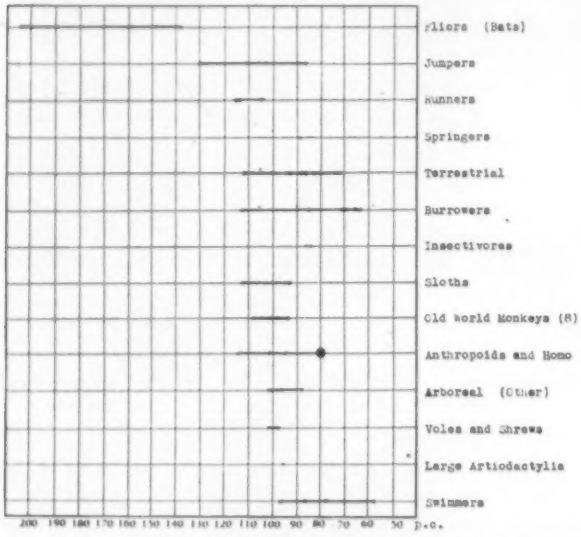


FIG. 38. Range of brachial index in different groups of mammals classified by function.

birth; this increases slightly to adulthood when it becomes 90. In the gibbon the index increases from birth, where it is 101, to adulthood when it is about 112 (Schultz, 1926: 490). It is a remarkable fact that while the two main arm segments grow before birth at nearly the same rate in man and anthropoids alike, after birth in man the rate of growth of the radius is immediately reduced while in the gibbon it continues for some months almost uninhibited.

This slowing up in growth of the human radius, as I pointed out (1937: 282) some years ago, "is probably not because of inability of the radius to grow as fast or for as long a time after birth as the humerus. It would seem rather that a new inhibiting gene has become active in the human infant to slow down the growth of the radius."

Though it has not been possible to discover the biochemical mechanism by which the radius slows down while the humerus continues to grow rapidly, the adaptive value of the process to man, as no longer a brachiating but now a tool-using organism, is obvious. Equipped as are man's upper limbs they are well suited for swimming, climbing, and the handling of tools. Probably the use followed the form.

Summary.—The brachial index tends to diminish with age, post partum; is generally less in girls than boys; is high in Negroes. Individual curves are wavy, indicating rhythms of growth. Still, curves are similar for monozygotic twins. Man's index of 80 places him close to the gorilla. In general, man's index fits him for swimming, mediocre climbing, and the handling of tools.

10. FOREARM (RADIUS) RELATIVE TO SHANK (TIBIA) RADIO-TIBIAL INDEX

Just as the upper and lower limbs are regarded by morphologists as "homologous" so the radius and tibia are also. In this section the differential growth of these limb segments will be considered. If the rate of growth of each remained uniform the percentage ratio would remain constant.

General.—Figure 39 shows the curve of change with age of this ratio. The prenatal part is computed from Scammon and Calkins (1929: 220, 243). Their landmarks are very different from mine, for from 6 fetuses nearing end of gestation I obtained the ratios indicated by the asterisks.

From 12 weeks to birth the percentage ratio diminishes about 6 points. Adjusted to my landmarks the change is approximately from 90 to 85. Thus during this period the tibia grows faster than the radius.

The mean value of the radius/tibia percentage ratio at birth in my series is about 85 for males and 83 for females. Kugler's data give a percentage ratio of 83.1 for males and 82.8 for females. According to Martin (1928: 429) Mollison found a ratio of 83.7 per cent for neonates.

It is of interest that the radio-tibial index undergoes greater changes with age than does the femoro-humeral index, inasmuch as the latter changes (in the male) only from 78 to 66 (12 points), as contrasted with the change from 85 to 65 (20 points) in the former. This is because of the slow growth of the forearm as compared with that of the shank.

Racial.—In my D2 series the Negroes, Italians, and Jews have a percentage ratio about a point greater than have the Nordics, that is, around 66 to 66.5.

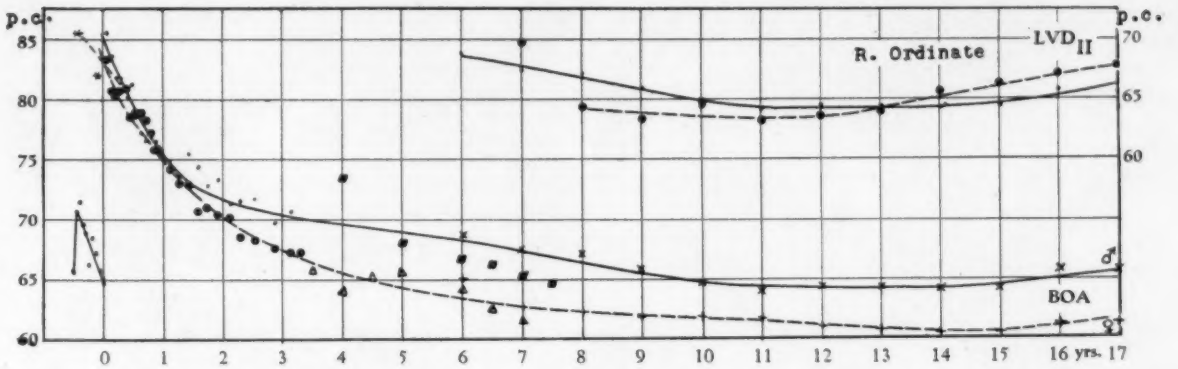


FIG. 39. Standard mass curves of change with age of percentage ratio of length of forearm to length of shank (radius/tibia). Fetal data from measurements by Scammon and Calkins on lateral aspect. Postnatal data from measurements on median aspect of limb. Ordinate scale for LVD series on right. The asterisks are based on my measurements of fetuses at the Department of Embryology, Carnegie Institution of Washington.

Comparative.—The radio-tibial index is in the apes over 100 (chimpanzee, 113; gorilla, 119; orang, 150; gibbon, 162). In the lower Primates it ranges from 84 to 108. It is a mere statement of fact that the human fetus starts with a high index, nearly like that of the adult chimpanzee. During development the tibia becomes relatively longer. This is an adaptive result since adult man is a walking, not a brachiating, animal.

We may extend the comparison to other mammals as given in table 7.

An examination of table 7 reveals that at the head of the list stand certain bats with indices of over 200, then follow certain tree-inhabiting animals—sloth, anthropoids, and Old World monkeys, with indices between 200 and 100. *Homo* with an index of 66 has lost his position among the other Primates since his radius has become so short. He takes his place now among the terrestrial mammals and swimmers; but hardly among the jumpers in which the index may be as low as 32. These last obviously must have a long and powerful tibia while the radius is semi-rudimentary.

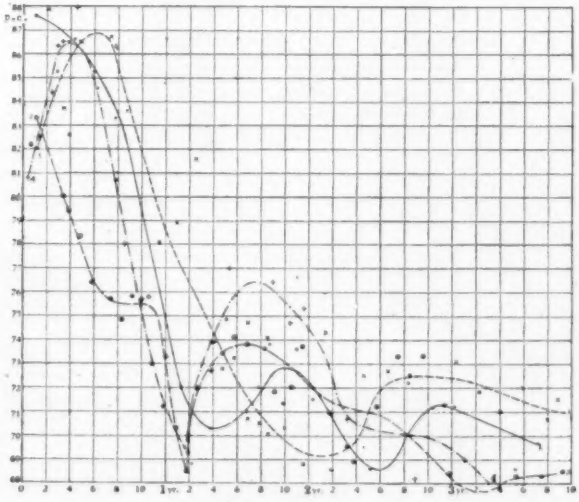


FIG. 40. Individual curves of change with age of percentage ratio of length of forearm to length of shank (radius/tibia) for 4 babies: 1, E. K., no. 1, f.; 2, C. C., no. 3, f.; 3, R. N., no. 6, f.; 4, T. P., no. 9, m.

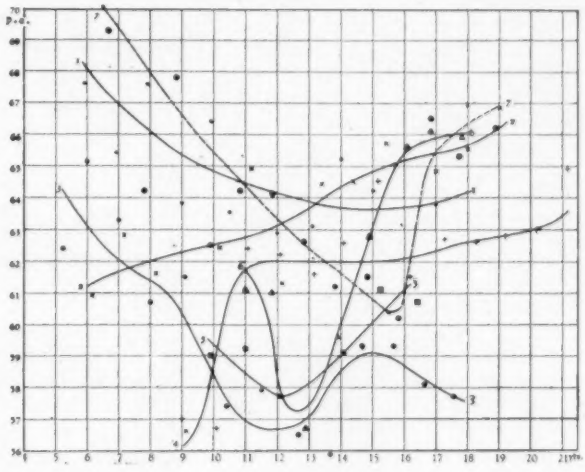


FIG. 41. Individual curves of change with age of percentage ratio of length of forearm to length of shank (radius/tibia) for the same 7 boys as in fig. 30.

Individual babies.—The radio-tibial index at or near birth (fig. 40) varies between 80.8 and 87.5. In some cases (nos. 3 and 4) the index increases from around 81 to 87, in other babies it falls rapidly to 69 at 14 months. When first measured standing the tibia shortens and the index rises to 74 or 76. Thereafter with some

TABLE 7

RADIO-TIBIAL INDEX (PERCENTAGE RATIO OF RADIUS LENGTH TO TIBIA LENGTH)

Eptesicus (bat)	253
Noctilio (bat)	240
Nycteris (bat)	230
Desmodus (bat)	223
Pteropus (bat)	210
Bradypus (sloth)	180
Symphalangus (169) (gibbon)	155
Hylobates conc. (gibbon)	150
Hylobates hool. (gibbon)	147
Orang (149)	147
Macacus nemes. (monkey)	146
Giraffe	121
Choloepus (sloth)	117
Elephas (elephant)	117
Gorilla (117)	116
Camelus (camel)	113
Chimpanzee (113)	110
Rhinoceros	105
Papio (baboon)	103
Ursus (bear)	103
Scalops (mole)	102
Cercopithecus pet. (monkey)	99
Macacus, var. sp. (monkey)	98
Cynopithecus (monkey)	97
Myrmecophagus (anteater)	97
Felis leo (lion)	95
Phascolomys (wombat)	94
Equus (horse)	94
Nasalis (monkey)	93
Canis (dog)	93
Erithizon epin. (porcupine)	93
Macacus leonius (monkey)	92
M. fuscatus (monkey)	93
Nycticebus (lemur)	91
Petaurus (flying phalanger)	89
Dasypodidae (armadillo)	89
Pygathrix (monkey)	88
Zalopus (seal)	88
Pithecus (monkey)	88
Cercopithecus (monkey)	86
Semnopithecus (monkey)	85
Dasyuridae (terrestrial marsupial)	84
Dendrolagus (tree kangaroo)	83
Sus (pig)	83
Cebus (monkey)	82
Semnopithecus (monkey)	82
Bovidae	82
Cerocebus (monkey)	82
Didelphus (opossum)	81
Cercopithecus (monkey)	81
Vulpes (fox)	81
Tapir	79
Colobus (monkey)	78
Cervidae (deer)	78
Conepatus (southern skunk)	77
Geomys (gopher)	76
Erinaceus (hedgehog)	76
Arctomis (woodchuck)	74
Felis diordis (cat)	74
Hemicentetes (insectivore)	73
Lepus (hare)	73
Obobaenus (walrus)	71

TABLE 7—*Continued*

Acanthion (porcupine)	71
Haplodon (sewellel)	69
Tayassu (peccary)	67
Eliomys (dormouse)	66
Solenodon (insectivore)	66
Homo (man)	66
Microtus (vole)	65
Hydrochoerus (water cavy)	62
Castor (beaver)	62
Thryonomys (burrowing African rodent)	61
Napæozapus (jumping mouse)	61
Sylvilagus (hare)	61
Potamogale (vole)	60
Manis (scaly anteater)	58
Chimarrogale (insectivore)	53
Phoca (seal)	52
Dipodomys (jumping mouse)	50
Zapus (jumping mouse)	48
Macropus (kangaroo)	48
Dipodomus (jumping mouse)	46
Lepus (hare)	46
Onychogale (kangaroo)	40
Macropus (kangaroo)	37
Pedetes (jumping hare)	35
Scirtopoda (jumping mouse)	32
Alactaga (jumping mouse)	32

fluctuations the curves fall to 67 or 72 as the tibia gradually assumes ascendency. The minimum at 2:6 years is hard to understand; perhaps it is due to a reaction of the tibia as it resumes its hereditary length.

Individual youth.—The curve of change of index with age between 6 and 19 years is shown in figure 41. In general the curves slope downward until about 12 years, at the time of the juvenile spurt. They thereupon ascend rapidly, as though the growth of the radius were more affected by this episode than that of the tibia. This corresponds to the rise in the index noted in the mass standard curve. Curve no. 6 is of a colored boy. No. 7 is of an idiot boy;

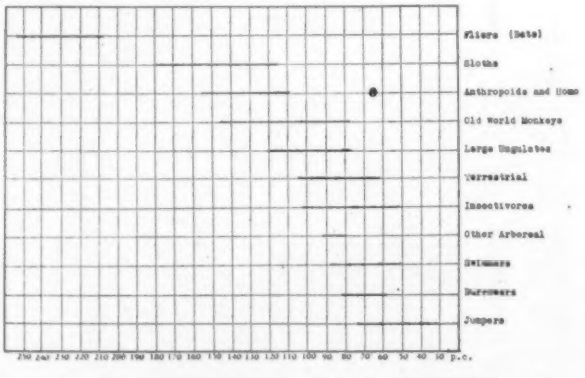


FIG. 42. Range of radio-tibial index (radius/tibia) in different groups of mammals classified by function.

at 18 years he has achieved the highest index of all.

Summary.—The index between the distal segments of the limbs is one between distantly "homologous" parts. In general the tibia grows faster than the radius, and the difference in this respect is most marked in girls. Babies show wave-like curves, as indeed do also juveniles and adolescents. The adult index is about 66. This places *Homo* far below the gibbon and orang, which have indices of about 147, and even below the gorilla (116) and chimpanzee (110). See figure 42. It places man functionally among the swimmers and burrowers. In this ratio man has quite lost the brachiating adaptation. In monozygotic twins the index changes with age in parallel fashion.

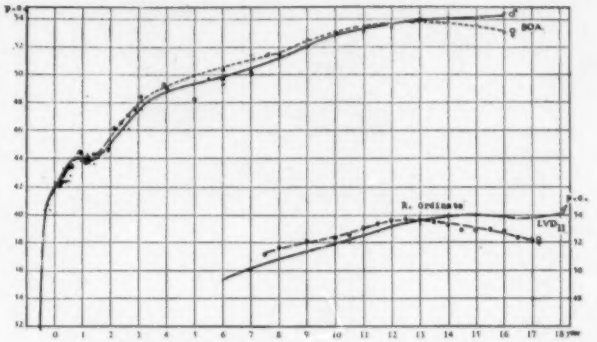


FIG. 43. Standard mass curves of change with age of percentage ratio of length of gross lower limb to stature. Ordinate scale at right applies to LVD 2.

11. GROSS LOWER LIMB RELATIVE TO STATURE

Data on this matter are afforded by figure 43, which shows the mean developmental curve of this relation.

Sexual.—Almost from birth the relative lower limb length is higher in preadolescent girls than in boys; the lower limb forms a larger proportion of stature. Indeed, as we have seen (sect. II, 5) the lower limb is absolutely longer in girls *until* 13 or 14 years of age. In the present curves a decussation occurs at 12 years (13 years in the D2 series). The total stature of course makes greater increase in length than the lower limbs

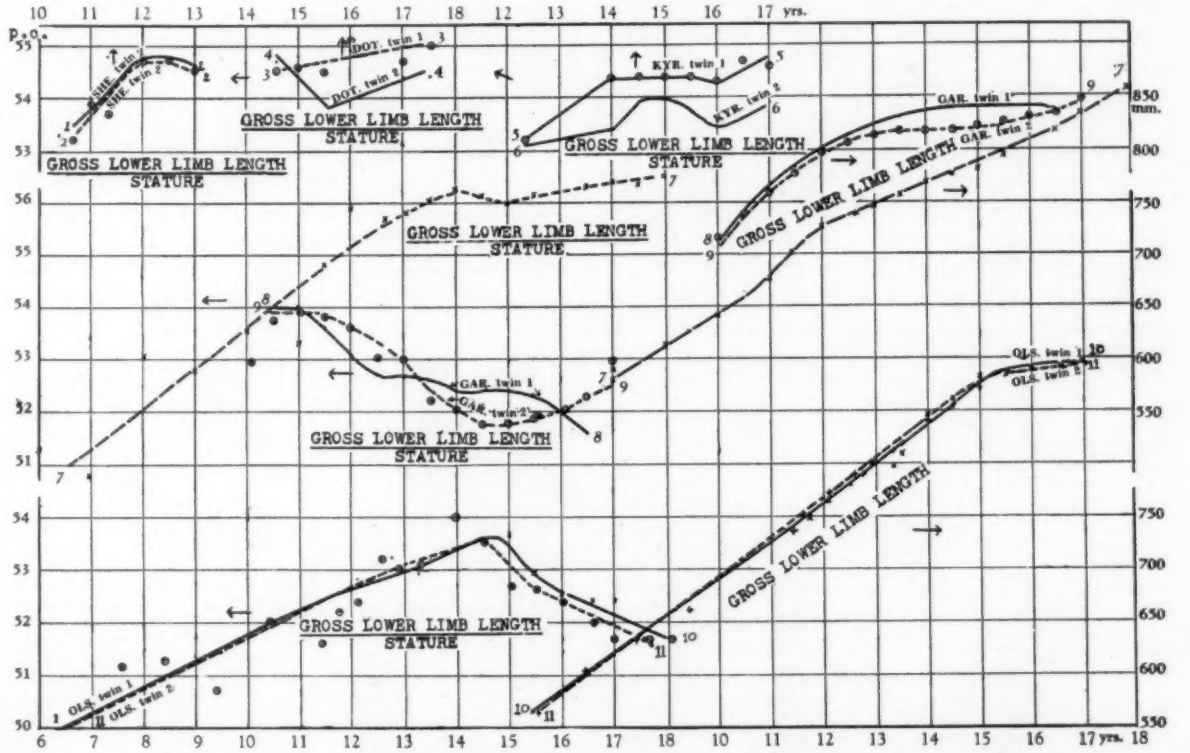


FIG. 44. On left individual curves of change with age of percentage ratio of gross lower limb length to stature for 5 pairs of twins and, for comparison, individual curves of gross lower limb length for two pairs of twins. Also growth in length of gross lower limb for no. 7, brother of GAR. twins. Left hand ordinates percentage ratio, right ordinates absolute measurements in millimeters. Vertical arrows point to proper scale of ordinates.

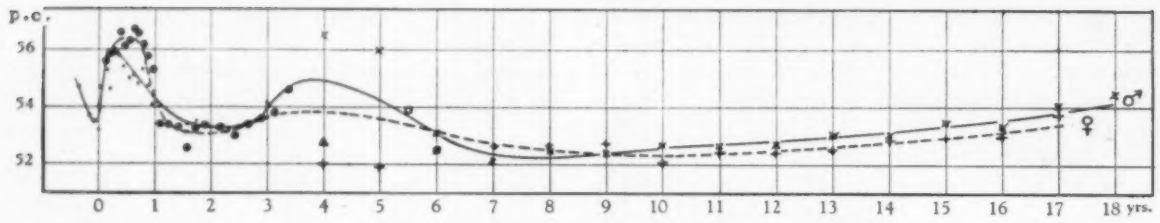


FIG. 45. Standard mass curves of change with age of percentage ratio of thigh length to net lower limb length.

alone do; and the relative lower limb percentage ratio falls in the mean girl from 54 to 53, or even 52, per cent of stature. A smaller relative limb length in adult females than in males is found on the average in some peoples of Europe, e. g. Belgians, but especially of Africa, e. g. Ashanti: m, 54.8, f, 54.0; Kru: m, 56.0, f. 52.2.

Figure 44 shows for certain twins the growth curves of gross lower limb length as well as the changes of the ratio of lower limb length to stature. It will be observed that whereas the gross lower limb length proceeds rapidly upward from 6 to 14 years, after 14 years the length of the lower limb ceases to increase with such speed. In fact the change in the case of the Ols. twins from 15 to 17 years is only 1 cm. p. a. Since the trunk continues to grow, relative limb length decreases about 2 points p. a. It will be observed that the curves of the ratio in the case of the Kyr., Ols., and She. twins lie rather close together. In the case of the Dot. Negro twins the curves though fairly close together are not at all parallel. Obviously the ratio is largely controlled by genetical factors.

Summary.—The percentage ratio of the lower limb's length to stature is increased from 42 at birth to 54 at maturity. Until 13 years it is larger in girls than boys; after 13 years smaller. It changes with age in parallel fashion in monozygotic twins.

12. THIGH LENGTH RELATIVE TO NET LOWER LIMB LENGTH

Length of the thigh and that of the shank ("leg" of anatomists) together constitute the whole of net lower limb length. The percentage ratio under discussion is that between the length of the thigh and the length of the net limb. Common observation shows that the thigh is the larger moiety.

As figure 45 shows, in the fetus of 20 weeks, the thigh region, which is the first part of the limb to develop, constitutes nearly 55 per cent of net lower limb. During the succeeding 20 weeks the percentage ratio falls to 53.5. After birth the thigh takes the initiative again and becomes about 56 per cent of the limb at 3 months p. p. This is indeed the peak of the thigh fraction. Within the following 21 months the percentage ratio of the length of the thigh to that of the net lower limb falls from 55.7 to 53.3. Its lowest position, at 52.2, is reached at 7 years. Thereafter, on the average, the percentage ratio increases to 54 at 18 years.

Though the data are few for the period $3\frac{1}{2}$ to $5\frac{1}{2}$ years it seems probable that there is a temporary maximum at about $3\frac{1}{2}$ years. This is in agreement with the findings for thigh length (fig. 20) which show a steep slope at 3 years while the slope of net lower limb is slightly reduced at this age (fig. 16).

The relatively long thigh is an anthropoid trait. Thus from data furnished by Schultz (1933) it appears that in the adult orang the percentage ratio of thigh to net lower limb is 53.8, in the chimpanzee 54.8, in *Gorilla gorilla* 54.5; in *G. beringei* 56.3.

Racial.—In my D2 series at 14 years the Negroes have the smallest index (53.3); the Nordics the largest (55.0).

Individual babies.—Figure 46 shows curves of change with age of the percentage ratio of thigh length to net lower limb length for four babies. In the four cases, at birth this has a median

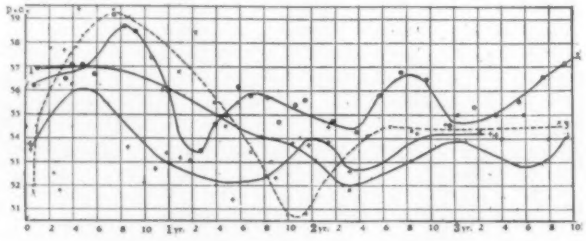


FIG. 46. Individual curves of change with age of percentage ratio of thigh length to net lower limb length for 4 babies: 1, E. K., no. 1, f.; 2, C. C., no. 3, f.; 3, R. N., no. 6, f.; 4, T. P., no. 9, m.

position of about 55. From 55 the ratio rises commonly about 2 points and in curve no. 3 it rises to over 59, all in the first 8 months. Thus the thigh grows relatively rapidly during earlier infancy as though released from intrauterine restraint. After that the ratio slopes downward in all cases and reaches a minimum in curve no. 2 at 14 months, in curve no. 4 at 18 months and in the other curves around 2 years. This downward slope, which in extreme cases reaches 51, indicates that, probably owing to the action of gravity after the child begins to stand, the femur shortens more than the tibia. After about 4 months from the minimum another maximum is reached, at various ages from 20 to 35 months. There are other changes of slope in the wavy lines which indicate that there is, as it were, a certain struggle between the femur and the tibia in the growth process so that now the one, now the other, takes the upper hand. The food supply to these two bones remains in all probability constant. There are alternating periods of most rapid assimilation and upbuilding of the bones with an interval between successive-like phases of about 6 months.

Individual youth.—As figure 47 shows, between 6 and 18 years the differentation curves of this ratio tend to slope steadily upward in some cases and to experience a minimum at the eleventh year in others. In some cases the upward slope is linear, in others wavy. The curve of the Mongoloid dwarf (no. 9) lies rather low. That of the acromegalic (no. 5) is of the wavy type. Curves nos. 6 and 7 are of two colored boys; in both cases the slope continues uniformly upward to 13 or to 18 years.

Summary.—The ratio of the length of the thigh to the net lower limb length is a variable one. The linear growth of the thigh seems somewhat depressed in utero, expands rapidly after birth, then slows up to a minimum ratio at 2 years. The curve thereupon rises and then shortly descends to mid-adolescence to increase slightly thereafter. The index is small in Mongoloid dwarfs and in Negroes. Individuals show much fluctuation in the ratio which is a delicate balance between length of thigh and of the rest of the limb.

13. SHANK (TIBIA) RELATIVE TO THIGH (CRURAL INDEX)

Eight years ago I published (1933) on this topic. Since then additional data have been obtained. These will permit a presentation of some new points while the general discussion may be curtailed. It may be repeated, however, that the Platyrrhine monkeys tend to have a higher index than adult man; among the Catarrhines the orang alone has a slightly higher index.

The standard mass curve is given for this ratio in figure 48. The curve starts prenatally at 12 weeks with a percentage ratio of 80. It then rises rapidly (following data afforded by Scammon and Calkins for white fetuses) until 84 per cent is reached at term. The values of the crural index found for neonates by different investigators are quite diverse. Thus Kugler found for males a mean value for the crural index of 93.2, and 94.5 for females. The mean in my series is 83.4 for males and 79.5 for females and the range of variation is from 100 to 63.5.

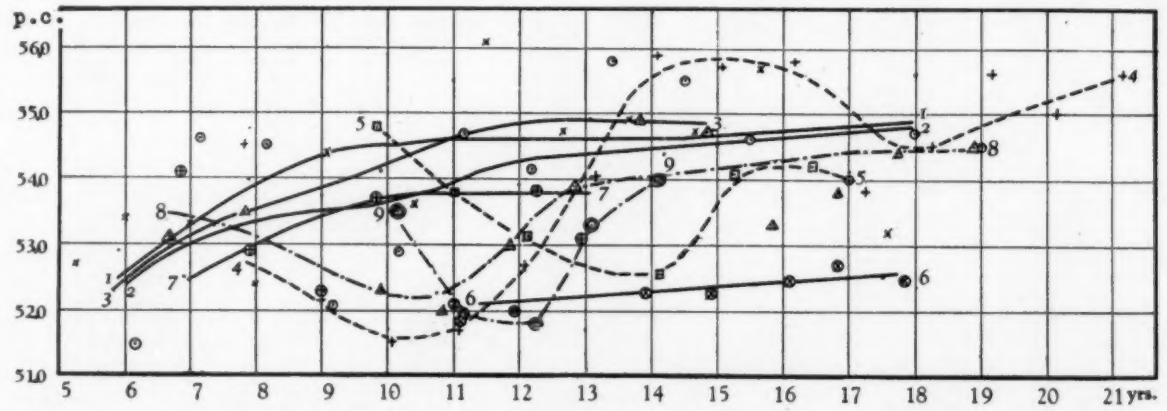


FIG. 47. Individual curves of change with age of percentage ratio of thigh length to net lower limb length for 9 boys: 1, J. G., no. 114 (Italian); 2, E. H., no. 1; 3, C. H., no. 83; 4, C. D., no. 30; 5, M. G., no. 95 (acromegalic); 6, R. H., no. 80 (col.); 7, H. T., no. 3 (col.); 8, J. M., I, no. 129; 9, S. C. (Mong.), no. 1.

For chimpanzees at birth Schultz (1940: 23) finds a crural index of 82.2 (80 to 86), as compared with 81.8 for "5 chimpanzee fetuses" (Schultz, 1933: 68). From immediately after birth in babies the percentage ratio decreases, reaching about 79 at 6 months. This is a period of more rapid growth in length of the thigh than

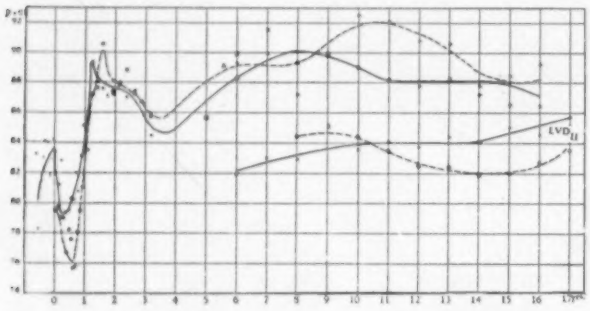


FIG. 48. Standard mass curves of change with age of percentage ratio of shank length to thigh length.

of shank (*cf.* fig. 22). During the middle of the second year the crural index increases rapidly reaching in the male over 89 as compared with 83 obtained by Schultz (1937: 302) from bones. In the infant chimpanzees measured by Schultz (1940: 23) the percentage ratio rises to 83.8. At about the fifteenth month the crural index of the human infant reaches a maximum which would seem to indicate that as the child first begins to stand on its lower limbs (at which time foot-vertex height may be temporarily even reduced) the limb bones are shortened, the femur more than the tibia. That is, there is a greater shortening by gravity of the femur as compared with that of the tibia during the period of learning to walk. This shortening is a temporary process and the femur, now strengthened, begins to grow rapidly again and so continues during the remainder of the year. Possibly the tibia gets the upper hand for a while at about 2 years, but after that the growth of the femur is the more rapid into the fourth year p. p. Thereafter there are alternating periods of less rapid and more rapid femoral growth. Especially at adolescence the spurt of growth in stature is due to rapid growth of the femur. After the adolescent spurt, growth of the two limb bones is of course reduced and to a similar degree. In the D2 series the mean curve of this ratio lies much below the standard. The boys are in general shorter and it appears that of the limb bones the tibia is the more reduced.

Social.—Of the groups measured by me at institutions the B. O. A. and D1 groups had, between 8 and 16 years, the lowest ratios (relatively shorter tibiae). The Mongoloid dwarfs after 14 years had the highest ratios (relatively long tibiae and short femurs). The absolute femur length of the Mongoloid dwarfs is strikingly shorter than that of the idiots or of the D series. The absolute length of the tibia of the Mongoloids is less strikingly smaller than that of the idiots. The lower the social status based on intelligence the later in life occurs the maximum index, i. e., the time of the longest tibiae.

Racial.—Of the groups measured, the Negro boys have strikingly the highest ratio; i. e., they have the proportionally longest tibiae. This agrees with Schultz' (1937: 302) findings from the skeleton, but not with those of Todd and Lindala (1928). The Mediterranean group has the lowest ratio (after 10 years). The absolute length of the tibiae as measured directly reveals the marked superiority in length of the Negro tibia.

Individual babies.—Figure 49 shows the change with age of the percentage ratio of the shank length to the thigh length for four babies. As

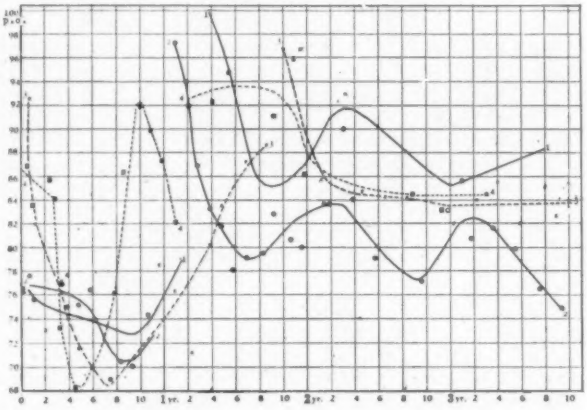


FIG. 49. Individual curves of change with age of percentage ratio of shank length to thigh length for 4 babies: 1, E. K., no. 1, f.; 2, C. C., no. 3, f.; 3, R. N., no. 6, f.; 4, T. P., no. 9, m.

we have already seen the femur is the larger, hence the percentage ratio will be generally less than 100. Of the four curves given, the median position at birth is about 82, although the four cases vary greatly in this index, namely from 76.5 to 92. For the first 5 to 9 months p. p. the curves slope downward (period of elongation of thigh) but thereafter there is a vigorous up-

ward slope for a while indicating a tendency of the tibia to win in the race with the femur. However, at about 13 months there is a great break in the continuity of the curves owing to the action of gravity in the standing child which results, as already seen, in the shortening of the thigh more than the shank. As a consequence the crural index takes a great leap upward in the scale so that in the case of no. 1 the value of 100 is reached. During the next few months as the femur grows stronger it straightens out and again leads the tibia in growth. However, in some of the curves around the second year a new advance takes place in the length of the tibia and the growth race between the two bones continues.

Individual youth.—Figure 50 gives the curves of change with age of the percentage ratio of shank length to thigh length for seven L. V. boys.

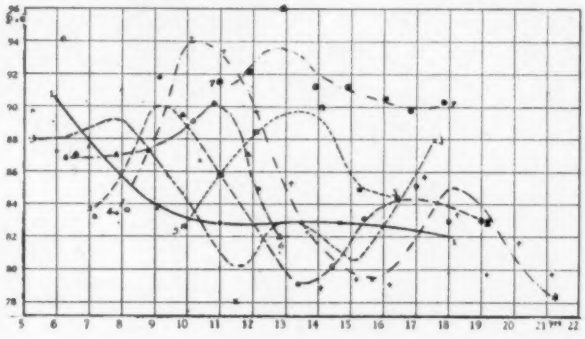


FIG. 50. Individual curves of change with age of percentage ratio of shank length to thigh length for 7 boys: 1, J. G., no. 114 (Italian); 2, E. H., no. 1; 3, C. H., no. 83; 4, C. D., no. 30; 5, M. G., no. 95 (acromegalic); 6, J. M., I, no. 129; 7, R. H., no. 80 (col.).

Most of the curves slope downward from 6 to 10 years onward. The period from 11 to 12 years is one of slight speed of growth of the tibia (+5 mm. p. a.) with rapid growth in the femur, 25 mm. p. a. During the period of the adolescent spurt (13 to 15 years) the tibia grows, on the average, 17 mm. p. a.; the femur 21 mm. p. a. Whatever stimulates growth at this time causes a greater addition to length in the femur than in the tibia. However, no. 5, that of an acromegalic, advances to 13 years; no. 2 reaches a minimum at 13½ years, and the increase thereafter is considerable. These curves picture again the relatively more rapid growth in length of the femur than the tibia during juvenility and adolescence. No. 6, of the idiot series, lies low in the series and shows a slight convexity upward. The curve of no. 7, the colored boy, has the highest position in this series, indicating that in this case the tibia is relatively longer compared with the femur than in white children.

Twins and special cases.—Figure 51 gives the curves of change with age of the percentage ratio of shank length to thigh length in the case of some

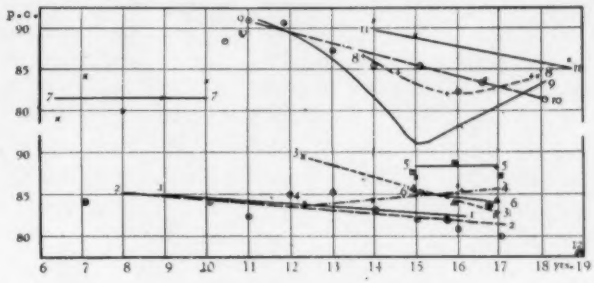


FIG. 51. Individual curves of change with age of percentage ratio of shank length to thigh length for twins and some special cases: 1, 2, A. G. and M. G. (monoz. twins), f.; 3, 4, M. K. and G. K. (monoz. twins), m.; 5, 6, E. D. and S. D. (dizygotic twins), f.; 7, A. A., cretin, f.; 8, C. B., treated cretin, f.; 9, M. C., atel., f.; 10, L. S., atel., f.; 11, M. W., atel., f.; 12, W. S., chondro., m., 29 yrs.

twins and special cases. Curves nos. 1 and 2 of the G. twins lie close together, hardly 0.1 point apart. Nos. 3 and 4, the Kyr. twins, show as in some earlier ratios a decussation, in this case at about 15½ years. Curves nos. 5 and 6 of the colored dizygotic Dot. twins remain from 2 to 5 points apart and they occupy relatively high positions owing to their race. No. 7 is the curve of an untreated cretin which occupies a relatively low position, whereas no. 8 of the treated cretin, is relatively high. No. 12, of the chondrodystrophic man, has a low position at 75. Curves of nos. 9 and 10, ateliotic dwarfs, lie rather low, that of no. 11, also an ateliotic dwarf, rather high. No. 9 is concave upward.

Summary.—The crural index increases rapidly during gestation until about 83 at birth is reached, but shortly after birth the femur grows so rapidly (as if earlier under restraint) as to force the index at 5 months down to 79. On standing, the femur seems to be shortened more than the tibia; but it soon recovers. Later a sort of rivalry ensues in the growth of these bones. The index is higher, on the whole, in girls than boys. Negroes have relatively long tibiae. In individual babies a sort of rhythm exists in the relative speed of growth of these bones. During the circum-adolescent period the

femur of white children tends to grow faster than the tibia. The ratio runs strikingly alike in twins.

14. SPAN RELATIVE TO STATURE

The dimension, span, formerly much stressed, is a complex of two arm lengths and shoulder width, with some over-lap. It is easily taken, and is of some interest in relation to stature (which in turn is largely made up of lower limb length which in turn is strongly correlated with upper limb length). In relation to stature span is generally nearly identical. I have touched on this dimension and this percentage ratio some years ago (1926: 217–220). One warning must be given in an attempt to compare the results of various anthropometrists. That is, the length of span is largely physiological-psychological, just as "grip" determined by the dynamometer is. The direction to "stretch" is given and the resulting length is supposed to be maximum. But the amount that a person can stretch is partly determined by his neuro-muscular control, by the vigor of reaction that he can command, by the tone of his muscles and by his attitude toward the anthropometrist. It has frequently been noticed that the lower grades of feeble-minded have a span much smaller than their arm length plus biacromion breadth would seem to call for. It is obvious on inspection that, however desirous the subject may be to co-operate, he cannot do so with the vigor of a person of stronger mentality.

As stated, span has often been determined and its relation to stature computed (Günther, 1938). A table of change in relative span with growth for seven racial stocks, and a table of relative span for groups of adults from nearly four score localities are given by Martin (1928: 405, 406) with percentage means running from 99 to 109. (See also a table by Wilder and Pfeiffer (1924: 544, 546).)

The control mass series of the percentage ratio is given in figure 52. The series begins at 1 year since it is hardly practicable to measure span on young infants. In the male there is a maximum at about 2 years, partly due to the temporary reduction of stature under the influence of gravity. At 3½ years a more normal relation seems to be established with a percentage ratio of 97.5. From that age the advance of the mean percentage ratio is steady. At 8 years 100 is reached and at 16 years 102.5.

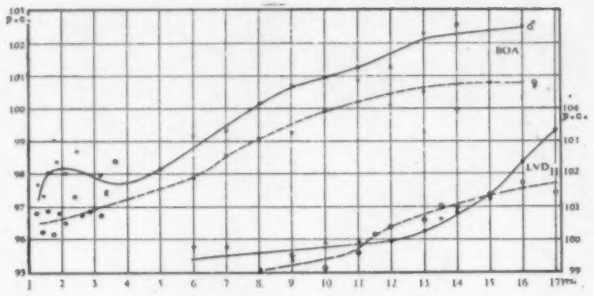


FIG. 52. Standard mass curves of change with age of percentage ratio of span to stature. Two sets of ordinates.

In the D2 series, boys at 17 years had indeed a percentage mean span relative to stature of 103.4.

Summary.—The physiological-psychological element in relative span leads to great diffuseness of determinations of points. There is a gradual rise of the percentage ratio from 97 at 3 years to 100 at 8 years to 102.5 at 16 years. Girls have generally a smaller index than boys—100.8 at 16 years. The control series has, at age 15 years, the largest index (102.5) and the Mongoloid dwarfs the smallest, 96.9. The percentage relative span of the American Negro boys was 107 or 5 points above that of Whites. In the long growth series a minimum commonly occurs in the thirteenth year. Chondrodystrophics have variable relative spans. An acromegalic has, at 16 years, a low value.

IV. CHANGE WITH AGE OF PERCENTAGE RATIOS OF GIRTHS

1. GIRTH OF "ARM" RELATIVE TO LENGTH OF UPPER LIMB

The general form of the upper limb is given by girth of the "arm" in relation to the length of the entire limb. The mass age changes are shown in figure 53. From data of Scammon and Calkins we may roughly compute the prenatal changes, which involve an increase in the percentage ratio of these two dimensions from about 37 at 12 weeks to 50 at birth. My data give 49 at birth; Kugler's (1932: 526) about 48 for boys. During the first 6 months p. p. the percentage ratio increases to over 54, because of the great increase in girth of the "arm." At the end of the first half-year a sudden change occurs. Girth increase slacks up, while (doubtless associated with that fact) gross arm length continues at a rapid rate. During the next 9 years the index has fallen on the average to 32.4 and

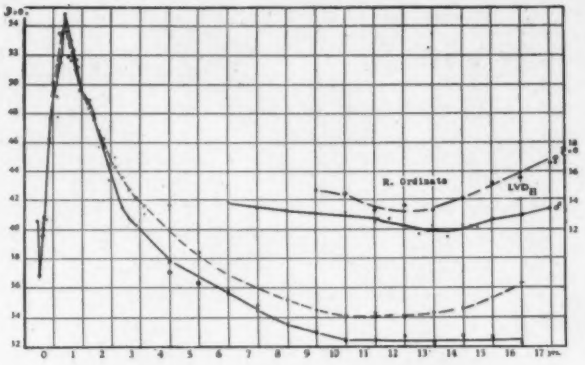


FIG. 53. Standard mass curves of change with age of percentage ratio of girth of "arm" to length of gross upper limb. Two sets of ordinates.

maintains nearly this relation to 16 years; with a tendency to rise slightly thereafter.

Sexual.—There is no obvious sex difference until the second year p. p. when the percentage ratio of the average girl is greater than that of the average boy by nearly 2 points and so continues to adolescence. Beyond that the excess is 3 or more points in the period when the contours of the girl tend to become more rounded.

Individual babies.—Figure 54 gives the curves of change with age of the percentage ratio of the girth of the upper arm to the gross arm length

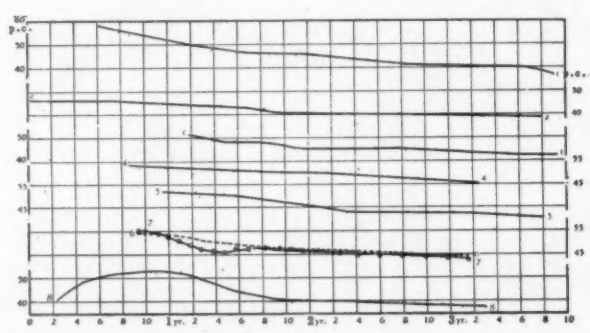


FIG. 54. Individual curves of change with age of percentage ratio of girth of "arm" to length of gross upper limb for 8 babies. Numbers as in fig. 29.

for 8 babies. At 1 year p. p. the percentage ratio lies at about 53, whereas at age 3 it has fallen to about 41. Obviously from birth onward, as the arm grows in length, it tends to become relatively slender. The curves of the twins, nos. 6 and 7, are extremely close together. The curves shown in the figure vary in slope. Thus in no. 1 the change in the ratio during the period from 6 months to 1:6 years p. p. is 12 points, whereas in no. 2, for the same period, it is 3 points. In general, after having reached 2:6 years the downward slope of the curves is diminished.

Individual youth.—Figure 55 gives the curves of change with age of the percentage ratio of girth of the "arm" to gross limb length for 6 boys. Before the age of 11 years the curves slope downward at the rate of about 1 point per year, but at about the thirteenth or sixteenth year some of the curves begin to move upward, becoming concave above. This result is seen especially in curves nos. 2 and 6 and perhaps

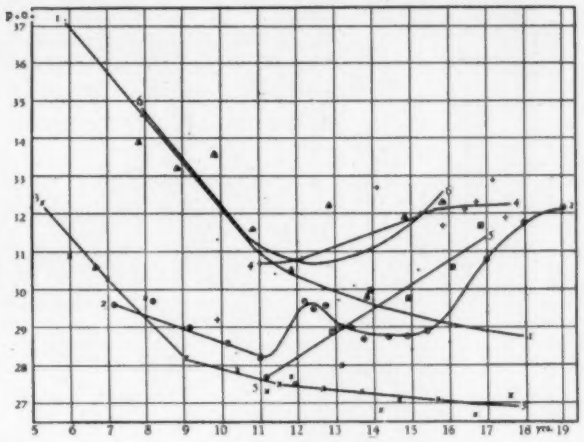


FIG. 55. Individual curves of change with age of percentage ratio of girth of "arm" to length of gross upper limb for 6 boys: 1, J. G., no. 114 (Italian); 2, E. H., no. 1; 3, C. H., no. 83; 4, M. G., no. 95 (acromegalic); 5, R. H., no. 80 (col.); 6, J. M., I, no. 129.

also curve no. 4. This suggests that during the adolescent and post-adolescent periods the girth of the "arm" increases relatively after the length of the upper extremity has ceased to increase much. This is probably due to the muscular activity of the children. However, in curves nos. 1 and 3 the upward slope has not yet occurred. Curve no. 5, that of a colored boy, proceeds upward from 11 to 17 years; there is no downward trend—perhaps because the observations began so late in life. Curve no. 6, for the idiot boy, J. M., occupies the upper part of the series, this despite the great arm length that is found among idiot boys in general and in J. M. in particular.

Twins and special cases.—Figure 56 gives the curves of change with age of the percentage ratio of girth of the "arm" to the gross arm length for some twins and special cases. Nos. 1 and 2 are

the curves of the G. twins. Their courses lie parallel and for the most part fairly close together, differing mostly about 0.5 point, although separated nearly 2 points at age 14. The curves 3 and 4 of the dizygotic Dot. twins are much more widely separated. The curve

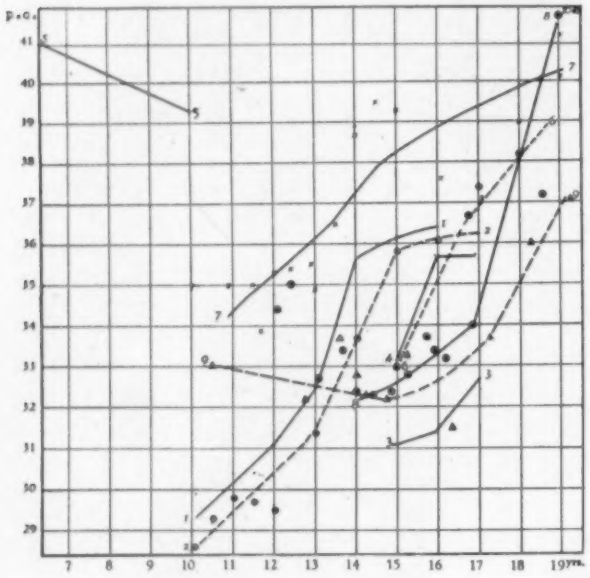


FIG. 56. Individual curves of change with age of percentage ratio of girth of "arm" to length of gross upper limb for twins and some special cases: 1, 2, A. G. and M. G. (monoz. twins), f.; 3, 4, E. D. and S. D. (dizygotic twins), f.; 5, A. A., cretin, f.; 6, C. B., treated cretin, f.; 7, M. C., atel., f.; 8, A. W., atel., f.; 9, V. R., microcephalic, m.; 10, W. S., chondro., m., aged 29 years.

of the cretin, no. 5, lies high up on the chart. The curve of no. 6, treated cretin, occupies a more median position. The chondrodystrophic man (no. 10) occupies a very high position, at 44.8 (outside the limits of the figure). He is a most typical chondrodystrophic dwarf, of age 29 years. It seems as though many of the curves of this chart are moving toward that elevated position of 44. No. 8, the curve of the ateliotic, has, indeed, almost reached that point at 19 years.

Summary.—The control curve is largely determined by the relatively slow growth in length of the upper limb before birth and by its rapid growth for 8 or 9 years thereafter. The age change in the ratio for babies shows chiefly a downward slope. The individual curves of boys run roughly parallel courses during 5 to 19 years, and are mostly concave above. In twins the differential curves run close. In dwarfs of some types the ratio is large.

V. DEVELOPMENT OF THE HAND

1. GROWTH OF HAND IN LENGTH

As someone has said, the hand of man is as extraordinary a human trait as the head. It is an organ of touch, of form and weight perception, of fine manipulation, of expression of the emotions. It has largely exchanged its ancestral tree-clinging function for a tool-using function. The tracing of its ontogenetic development should prove interesting.

Method.—The technique of measuring the hand of the living is so different in the practice of different anthropometrists that their results are not closely comparable. An analysis of the varied techniques has been made by Bayer and Gray (1933). My technique is described in my little guide (1927: 39). An outline of the hand is drawn while the hand and arm lie horizontally on the table, pronated, and with the hand axis in continuation of the forearm axis. The contour is traced with a split pencil and measurements made from it. For length the proximal point is the mid-point on the line joining the tips of the styloid processes; the distal point the tip of the third digit. For width the distance between the most lateral parts of the heads of the second and fifth metacarpals was measured. The hand lengths as measured by me are 1 to 3 mm. greater than those of Hrdlička, who measures from the distal wrist wrinkle. They make the sum of hand, forearm and arm practically equal to the length of the entire limb. Of course many other dimensions of the hand might be measured as listed by Wechsler (1939: 207, 211). My findings are summarized in the figures 57 to 61 and accompanying text. Length and width were the only measurements made by me.

In the early measurements (B. O. A. series) the hand length was measured in a different way; namely, with the subject standing erect, the distance was measured with the anthropometer from the floor to the apex of the middle finger and also to the tip of the radial stylion. The difference between these measurements was taken as hand length. This method, which was used in getting the control curve in these young people, gives results about 5 mm. shorter than that obtained from the tracing of the hand.

Figure 57 gives the curve of mean growth of the hand length from mid-gestation to late puberty. Growth is extremely rapid before birth, at the rate of about 90 mm. p. a. At birth the hand length in my Nordic series was about 62 mm. This compares with 64 mm. in the case of males found by Weissenberg for South Russian Jews, by Daffner (1902: 438) for Leipzigers and by Kugler (1932: 477) for Zurichers, and with 59 mm. found by measurements made on photo-

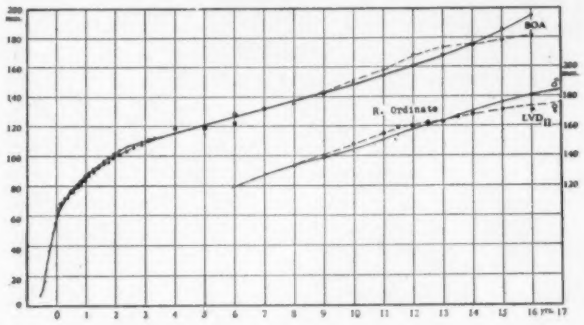


FIG. 57. Standard mass curves of growth of hand in length. Two sets of ordinates.

graphs of both sexes by Friedenthal (1914: 129, 130). Thus the human hand of West European stock at birth is a little over 6 cm. long.

During the first postnatal year the mean hand length increases 27 mm.; during the second year, 16 mm.; and during the third, about 9 mm.; thus absolute growth steadily declines until at the sixth year it is only 6 mm. p. a. There is a rapid increase in hand length during the fifteenth year which is that of mean adolescent spurt of growth. In that year growth is 15 mm. p. a. or about the same as during the second year. Meanwhile the hand has increased greatly in length, so the percentage increase is small. But percentage increase does not mean much where growth is strictly terminal.

The increase of the hand in length is the summation of growth of the various carpal bones and (especially) of the metacarpals and first, second and third phalanges. During rapid growth all these bones are growing in length with about equal rapidity. Take for example the hands of Katherine Kl. and Rachel Le. I give for each girl in Table 8, first line, the age, initial lengths in mm. of the third metacarpal and of the proximal and middle phalanges measured from radiograms; and below these data (on the second to fourth lines) the increments in millimeters to the initial length in successive years of ages.

TABLE 8

INITIAL LENGTH AND SUBSEQUENT INCREMENTS IN LENGTH OF BONES OF THE HAND

1. Kath. Kl.

Age	Right Hand			Left Hand		
	3rd Metacarpal	Prox. Phal.	Mid. Phal.	3rd Metacarpal	Prox. Phal.	Mid. Phal.
11	59	43	24	58	42	24
12	2	2	1	2	2	1
13	5	4	2	4	5	3
14	7	6	4	6	7	4
	2. Rach. Le.					
9	38	28	17	38	29	18
10	2	2	1	3	1	0
11	4	3	2	4	3	0
12	6	5	3	6	4	2

In both cases the third metacarpal grows fastest, increasing in length about 2 to 3 mm. per year; the proximal phalanx about 0 to 2 mm. slower, and the median phalanx about 1 to 3 mm. slower still than the proximal phalanx.

The growth of hand length of the males slows down greatly after the adolescent spurt and in the B. O. A. series probably becomes asymptotic at about 20 cm. Hrdlička (1925) finds mean left hand length for adult male Old Americans to be, in his series, 19.3 cm., but he measured to a different base line. Compare: Schlaginhaufen (1932), Swiss recruits 19.1 cm.; Wechsler (1939: 218), Northeast Swiss, males 19.0 cm., females 17.8 cm.

In the D2 curve, shown in the lower part of the figure, the mean males at 17 years reached a hand length of 185 mm. in accordance with the prevailingly short stature of this series. The curve would probably become asymptotic below 200 mm.

Sexual.—During the first postnatal year the mean hand length of the girls runs about 2 mm. shorter than that of boys, and this difference is increased about a millimeter during the third year. After that the curves of the sexes approach for a time and decussate at about 7 years. Thereafter, until the period of the male adolescent spurt, the girls' hands are about 3 mm. longer than those of boys. Growth in length nearly stops, at 16 years, at 182 mm.

In the D2 series the relative position of the curves of the sexes is about the same as in the

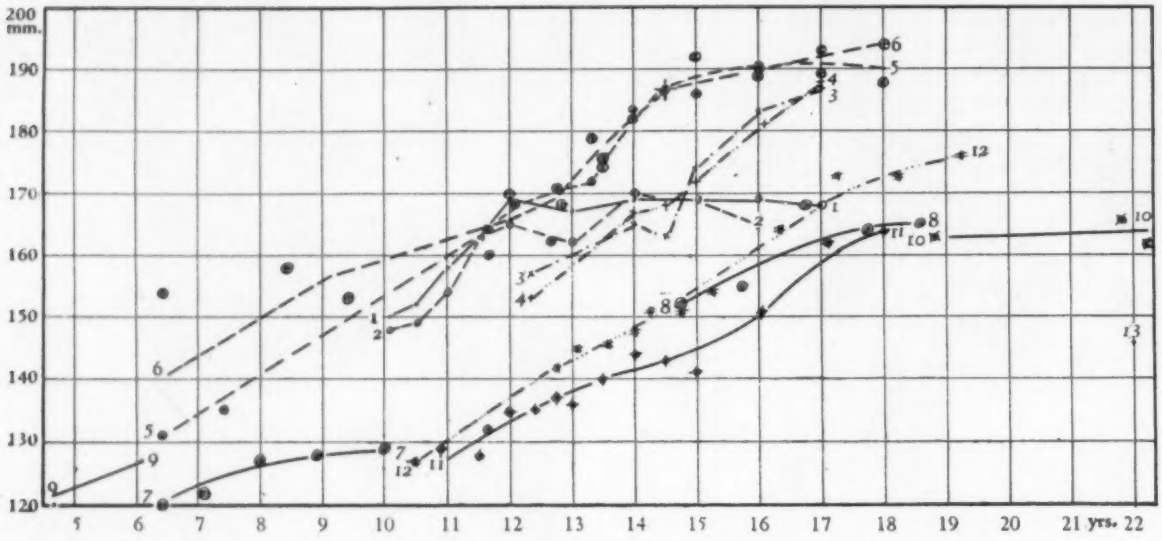


FIG. 58. Individual curves of growth of hand in length for twins and some special cases: 1, 2, A. G. and M. G. (monoz. twins), f.; 3, 4, M. K. and G. K. (monoz. twins), m.; 5, 6, Ri. O. and Ro. O. (monoz. twins), m.; 7, A. A., cretin, f.; 8, C. B., treated cretin, f.; 9, H. L., cretin, m.; 10, E. S., chondro., m.; 11, M. C., atel., f.; 12, V. R., microcephalic, m.; 13, W. S., chondro., m., aged 29 years.

control. The second decussation occurs at the fourteenth year and growth practically stops, with girls, at 16 years at about 175 mm.

Social.—Of the four groups considered, the control group has the largest hands as it has the greatest body size—at least this superiority begins to show itself at adolescence. This superiority in hand length of the more fully developed individuals is a trait that often first shows itself at, and after, adolescence. The eventual great stature seems to be a resultant of excessive activity of the growth-promoting hormones in large boys. On the other hand, the Mongoloid dwarfs have the shortest hands of all, as they have the shortest main limb bones.

Racial.—The American Negro has, on the average, the longest hands, in agreement with Virchow's findings (1879), which have been repeatedly confirmed (*cf.* Todd and Lindala, 1928). The mean Nordics and Mediterraneans have hands of similar length during pre-adolescent life. After adolescence the hand length of the Nordics becomes superior.

Twins and special cases.—In identical twins the growth of hand length is often extraordinarily alike. Thus (fig. 58) nos. 1 and 2 are curves of the G. twins. They lie nearly parallel throughout. Nos. 5 and 6 are curves of the Ols. twins. They are alike after 11 years (Ri.'s curve is based on only two points before that age). Curves nos. 3 and 4 could hardly be much closer.

Of the special cases, the dwarf, no. 11, has short hands, about 150 mm. long at 16 years. The chondrodystrophic, no. 10, has longer hands, about 163 mm. at 19 years, but no. 13 has a hand only 145 mm. long. The cretins, nos. 7 to 9, have short hands, about 125 mm. among the six-year-olds. It is rather remarkable that the chondrodystrophics with short extremities still have fairly long hands. The microcephalic, no. 12, has (considering his short stature) fairly long hands—176 mm. long at 19 years.

Summary.—Growth of the hand in length proceeds fairly smoothly at the rate of about 75 microns per day in first year babies; 25 microns per day in third year babies, and about 16 microns in adolescent boys. The growth curves of individual boys advance irregularly parallel between 7 and 18 years; but at very different levels. The hands of twins grow similarly in length.

2. GROWTH OF HAND IN WIDTH

Figure 59 shows the mean growth curve of the right hand from birth to maturity. At birth, the hand width in my series is 36 mm. in boys and 35 mm. in girls. Kugler (1932: 477) finds 39 and 38 mm. respectively. The slope of growth curves during the first year is 11 mm.

p. a.; during the second year 5 mm. p. a. After 3 years the mean growth of hand width is slow for a time and then increases at 5 to 7 years. An increase at 13 to 14 years in the D2 series is perhaps associated with increased use of the hands in labor such as with pick and shovel and in milking.

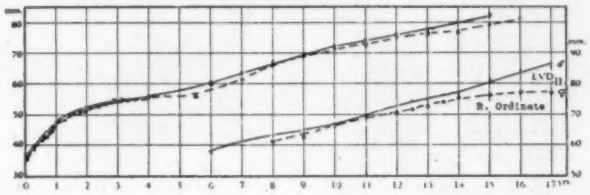


FIG. 59. Standard mass curves of growth of right hand in width. Two sets of ordinates.

Twins and special cases.—Figure 60 shows the growth in width of the hand in three pairs of twins. Nos. 1 and 2 are of the G. twins. These run somewhat parallel courses, about 1 mm. apart. Nos. 3 and 4 are of the Ols. twins; these also run nearly parallel courses, about 3 mm. apart, diverging for a brief period during adolescence by about 8 mm. Curves nos. 5 and 6 are of the Kyr. twins; they lie exceedingly close together, only 1 or 2 mm. apart. This figure then illustrates the closeness of control, even in hand width, of the genetic factors.

Figure 60 shows also the growth of hand width in some special cases. Nos. 7 and 8 are of hand width of cretins. Curve no. 9 is of the growth of hands of a chondrodystrophic girl who attains a hand width of 79 mm. by 22 years. No. 12 is of a chondrodystrophic boy who at 29 years has a hand width of 77 mm., whereas 83 mm. is about normal. The hand has a small width in these dwarfish children. No. 11 is the curve of hand width in a dwarf girl. Her hand attained a size at 19 years of 72 mm., whereas the hand width of V. R. (microcephalic) at the same time is 82 mm.

Summary.—The hand grows in width slowly as compared with length. It seems to be more reactive to use than hand length. In twins there is a close resemblance in growth.

3. PERCENTAGE RATIO OF HAND WIDTH TO HAND LENGTH (HAND INDEX)

This is the best index of the form of the hand. This form is very variable in Primates from *Pongo* to *Homo*, as illustrated by Midlo (1934: 365). In general the arboreal species have the slenderest hands. We shall look with interest on the form changes with age in *Homo*.

Figure 61 gives the changes with age in the hand index from birth to maturity. Some data from fetuses given by Schultz (1926: 492) do not tie up well with my data from new-borns, probably owing to different techniques. Schultz finds that the index falls from 57.8 at 10 weeks to 52.7 at birth.

At birth the index in living babies as found by me is 57 for boys; 55 for girls. Daffner (1902: 438) found an index of 54.7; Kugler's (1932: 477) data give an index of 53.2 for boys and 53.5 for girls. The index declines rapidly to 49.4 at 3 years and 48 at 5 years in boys. Eventually, in the B. O. A. boys it reaches 46. Schlaginhaufen (1929) found an index of 46.8 for Swiss recruits. In the D2 boys the mean index decreases from 51.7 at 13 years to 47 at 17 years. The figure also gives, at the bottom, the curve of age-change of index for Viennese boys copied from Brezina and Lebzelter (1925: 386). This shows a minimum of 42.5 at 12 to 13 years with an increase to 44 by 18 years, indicating very long hands. But the authors themselves suggest this may be due to a peculiarity of their technique, though they do not state what their technique is.

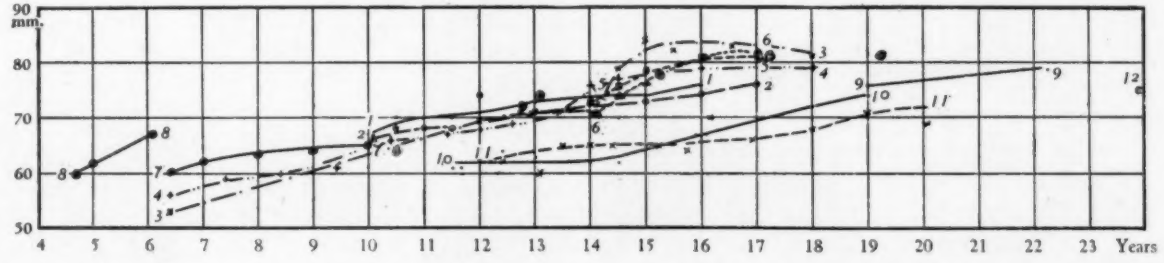


FIG. 60. Individual curves of growth of hand in width for twins and some special cases: 1, 2, A. G. and M. G. (monoz. twins), f.; 3, 4, Ri. O. and Ro. O. (monoz. twins), m.; 5, 6, M. K. and G. K. (monoz. twins), m; 7, A. A., cretin, f.; 8, H. L., cretin, m.; 9, E. S., chondro., f.; 10, M. C., atel., f.; 11, A. W., atel., f.; 12, W. S., chondro., m.

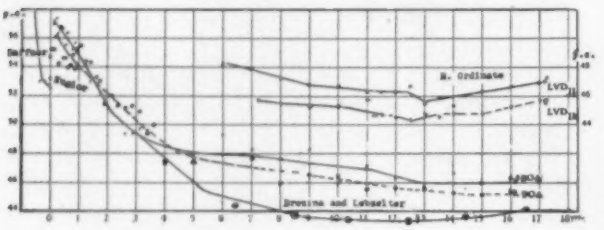


FIG. 61. Standard mass curves of change with age of percentage ratio of hand index (width/length). Two sets of ordinates.

Racial.—The Nordic hand has, after 10 years, a high index (47) and the American Negro a relatively smaller one (47 to 44.5) indicating long narrow hands (*cf.* Hrdlička, 1928: 25). This Nordic index is much the highest found among the Primates. The hand index is, according to Schultz (1926: 492), in the adult orang-utan 22 to 28; gibbon 22 to 25; howler monkey 31 to 34. The Nordic hand has progressed furthest in the direction of the compact short-fingered hand. Since the adult anthropoids have a small index one might hastily conclude, with Vallois (1932), that the baby's hand does not recapitulate. Actually the fetal anthropoids have broader hands (index 32 to 28 in the chimpanzee) than the adult anthropoids (Schultz, 1926: 492; Vallois, 1932: 915), though not as broad as has the human fetus. Fetal anthropoids anticipate the extraordinarily narrow hand of the adult. Here again the human and the anthropoid proportions travel the same road for a while in ontogeny; finally their paths separate to move toward their respective adaptive ends.

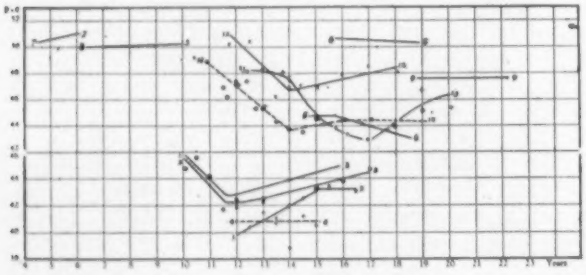


FIG. 62. Individual curves of change with age of percentage ratio of hand index (width/length) for twins and some special cases: 1, 2, A. G. and M. G. (monoz. twins), f.; 3, 4, Ri. O. and Ro. O. (monoz. twins), m.; 5, A. A., cretin, f.; 6, C. B., treated cretin, f.; 7, H. L., cretin, m.; 8, M. C., chondro., f.; 9, E. S., chondro., f.; 10, W. S., chondro., m., aged 29 years; 11, M. C., atel., f.; 12, L. S., atel., f.; 13, A. W., atel., f. Two sets of ordinates.

Twins and special cases.—Figure 62 gives the curves of change with age of the percentage ratio of hand width to hand length of twins and some special cases. At the bottom are shown the curves of the G. twins which run parallel courses at a distance apart ranging from 0.4 to 0.8 points; also of the Ols. twins, whose curves decussate. Among the special cases, the chondrodystrophics nos. 9, 10, and 11 have stubby hands. Of the cretins nos. 5 and 7 have fairly stubby hands, whereas no. 6, who received thyroid therapy, has more slender ones. The ateliotic dwarfs, nos. 11, 12, and 13, have hands of nearly normal proportions.

Summary.—The decrease in the hand index before and after birth is a characteristic of both anthropoids and man, as data of Schultz (1933: 68) suggest. Apparently from the youngest chimpanzee fetus (corresponding in age roughly to human fetuses of at least the end of the sixth lunar month) to the oldest (corresponding to the end of the eighth month) the index diminished from 45.6 to 45.1. The adult chimpanzee (average of seven) gives a percentage ratio of 34.7, very much smaller than in man.

VI. DEVELOPMENT OF THE FOOT

This topic was considered by me in a paper published in 1932. Consequently, although many additional data have been obtained since, the subject may now be considered in abbreviated fashion in an attempt to avoid repetition.

In introducing the subject it is desirable to recall that the European foot is of the plantigrade type with both longitudinal and transverse arches, the phylogenetic significance of which has been stressed by Morton (1922–24).

1. GROWTH OF FOOT IN LENGTH

Figure 63 gives the standard mass curve of growth in length of the foot from mid-gestation to maturity. From the tenth week to the fortieth the fetal foot increases in length from 8 mm. to 76 mm. At birth 18 male babies gave a mean length of 80 mm. and 15 females a mean of 83 mm. Kugler obtained a neonatal foot length of 80.1 mm. for boys and 78.9 mm. for girls. Weissenberg (1911: 68) finds a mean of 78 mm. for male neonates and Friedenthal (1914: 115) finds a mean of 63 mm. Thus the foot length is, at birth, one-third greater than hand length. The two organs retain about the same proportion of 3:4 to maturity; if anything, the hand length

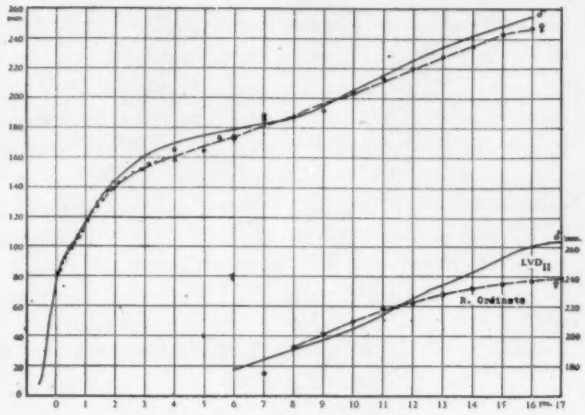


FIG. 63. Standard mass curves of growth in foot length. Two sets of ordinates.

falls a trifle behind the foot length. The increase is about 37 mm. in the first year, 26 mm. in the second, and 18 mm. in the third. Already growth-inhibiting factors are at work.

Sexual.—The mean length of the foot in girls is less than that of the boys except at 8 to 9 years (8 to 11 years in the D2 series).

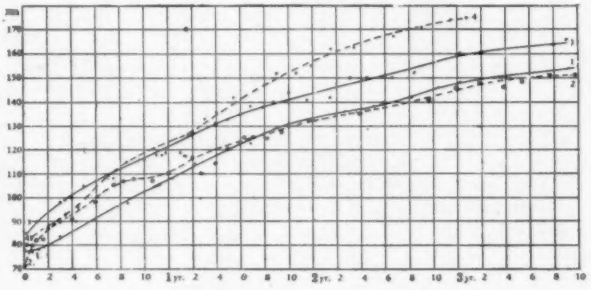


FIG. 64. Individual curves of growth of foot length for 4 babies: 1, E. K., no. 1, f.; 2, C. C., no. 2, f.; 3, R. N., no. 6, f.; 4, T. P., no. 9, m.

Individual babies.—Figure 64 gives growth curves of foot length for four babies. The boy (no. 4) at 10 days occupies a median position but at 10 months has the longest foot, far exceeding any of the girls. The curves show a few temporary fluctuations in foot length, some associated with standing.

Summary.—The growth of foot length proceeds rapidly before birth, more slowly immediately after, and slower still at 4 to 8 years. The growth curve of girls lies below that of boys. Except for a rapid increment in standing the curves are fairly uniform. At age 20, foot length ranges between 20 and 26.5 cm. in the special cases.

2. GROWTH OF FOOT IN WIDTH

This growth curve of foot width (fig. 65) is in striking contrast with that of foot length. The width, which at birth is only a little less than half the length, becomes, at 16 years, a little less than one-third of length. The width of foot in the neonate is 34.5 mm. for boys and 33.6 mm. for girls. Kugler (1932: 527, 537) found 32 and 31.1 mm. respectively for Swiss neonates.

Sexual.—The foot width of the girls is constantly less than that of the boys, though the curves approach closely at 8 years, just as the foot length curves do. In the D2 series the difference between the sexes is even greater than in the standard (B. O. A.) series.

Twins and Special Cases.—Figure 66 shows a great range of curves of foot width, from the G. twins (nos. 1 and 2) to that of the ateliotic dwarf (no. 9).

3. CHANGE WITH AGE OF FOOT INDEX

The foot index is the percentage ratio of width to length. Figure 67 shows that at birth the mean index is about 42.5 for males and 42 for females. From Kugler's (1932) means we compute the neonate indices for her babies to be 40.0 and 39.4. From birth the index diminishes to 36 at 16 years, because the length of the foot increases faster than the width.

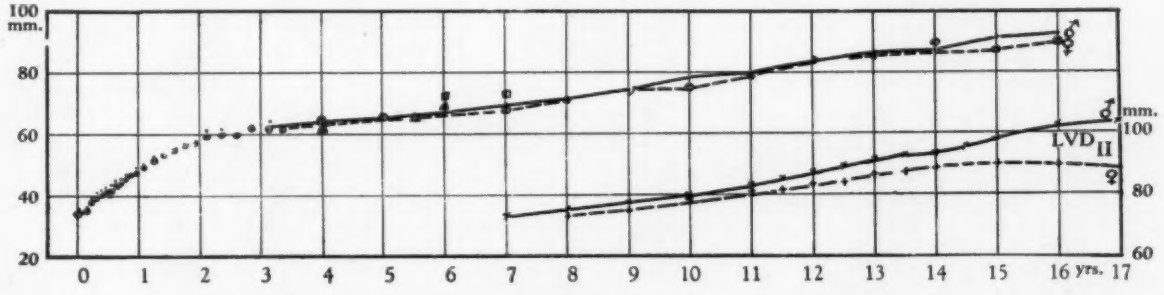


FIG. 65. Standard mass curves of growth of foot width. Two sets of ordinates.

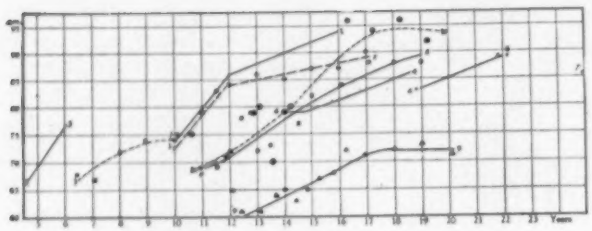


FIG. 66. Individual curves of growth of foot width for twins and some special cases: 1, 2, A. G. and M. G. (monoz. twins), f.; 3, A. A., cretin, f.; 4, C. B., treated cretin, f.; 5, H. L., cretin, m.; 6, E. S., chondro., f.; 7, W. S., chondro., m., aged 29 years; 8, M. C., atel., f.; 9, A. W., atel., f.; 10, V. R., microcephalic, m.

It appears that the foot index of the chimpanzees "Dayton" and "Evo," whose feet were repeatedly measured by Schultz (1940: 25), was about 47 at 4 years as compared with 37 for children of that age. Also in the chimpanzees the index *increases* with age to about 58 at 14 years. That is, the foot is getting relatively *broader*. It appears, then, that the foot index of the human and that of the chimpanzee are

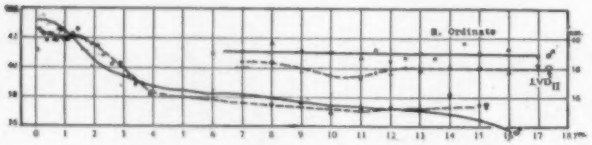


FIG. 67. Standard mass curves of change with age of foot index. Two sets of ordinates.

similar in infancy but, in subsequent development, diverge, the foot of the human child becoming relatively narrow. In the D2 series the indices are much larger than standard: 39 instead of 37 in adults.

Sexual.—In the control series the curve of foot index is the larger in the average girl in the second and third years but smaller between $3\frac{1}{2}$ and 12 years. Then a decussation occurs again. In the D2 series the index of the girls is uniformly below that of the boys. The wide feet of the D2 boys may be in part due to the rough work and rough shoes to which these boys are accustomed.

Social.—The foot index is smallest for the control series; greatest for the Mongoloids who in my series have an index of 44 at 18 years of age.

Racial.—In the colored (Negro) boys the foot index is nearly a point higher than in the Whites, after 11 years of age as Hrdlička (1928: 25) finds for full blood Negroes. Girls show a less marked excess.

VII. INTERRELATION OF CERTAIN DIMENSIONS

1. INTERVAL BETWEEN HEIGHT OF RADIALE-STYLION AND TIBIALE HEIGHT

As the arms hang by the sides of persons standing erect a great variation is noticeable in the distance between finger tip and knee. Our data give some means of expressing this distance quantitatively. Since the finger tip is not an accurate landmark (chiefly on account of a tendency of the fingers to be partly flexed) stylion and tibiale are taken as points of reference. Our problem then is how variable is this interval in the grown person and how does it change with age? This distance has some interest because of the fact that in anthropoid apes the wrist hangs much lower with reference to the knee than in humans and because it is popularly thought that the arms hang low in the case of low grade feeble-minded children.

Figure 68 shows the mean age change in the distance between stylion and tibiale, both measured from the floor. At the third year p. p. when the thigh is elongating rapidly (fig. 20)

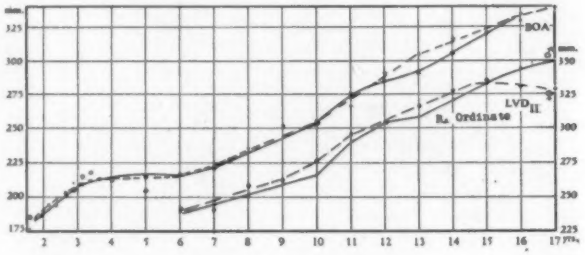


FIG. 68. Standard mass curves of change with age of interval between stylion and tibiale. Two sets of ordinates.

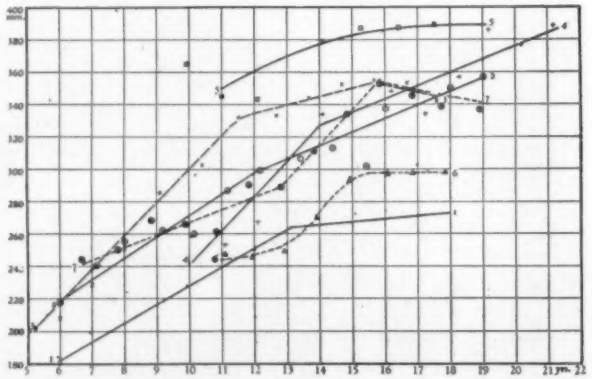


FIG. 69. Individual curves of change with age of interval between stylion and tibiale for 7 boys. Numbers as in fig. 30.

the distance increases rapidly and it remains nearly stationary in the fourth to sixth year when growth of the thigh has slowed down. At the period of the adolescent spurt in boys the interval increases greatly. The curve of the D2 series runs parallel to the standard.

Sexual.—The mean curve of the girls runs close to that of the boys except at the time of the adolescent spurt which is about 2 years precocious in girls.

Taking 13 boys and 18 girls of the B. O. A. series at the sixteenth year, it appears that this distance falls into the following classes with the following frequencies:

Class	26± 1 cm.	28 ±1	30 ±1	32	34	36	38	Mean cm. above Tibiale
13 m.			3	4	3	2	1	32.31
18 f.	1	0	5	5	2	3	2	32.67

If we take 17.0 cm. as hand length then in the 28.0 cm. class the finger tip will be only 11 cm. above the knee. In the 38 cm. class it will be 21 cm. above the knee.

In the D2 series we have the following distribution of intervals at 12 and 14 years respectively for each sex.

Class	26± 1 cm.	28	30	32	34	36	38	Sum	Mean cm.	St. Error
12 years m.	1	5	8	9				23	30.2	∓0.3
f.	2	5	11	8	1	1	1	29	30.5	∓0.3
14 years m.		1	6	10	6	2		25	32.2	∓0.3
f.		2	7	7	10	4	1	31	32.7	∓0.3

It appears that there is no significant difference between the sexes in this dimension in either the standard or D2 series.

Examples of the different intervals are illustreated in plate 1, *C* which shows a long interval (343 mm. at 15 years); *D* shows a short interval (294 mm. at 13 years).

Individual babies.—In all cases the curves slope upward, indicating that the distance between knee and wrist is increasing with age between second and fourth years. In a few cases the advance is irregular indicating that there is a difference in the rate of growth of the long bones of upper and lower extremities at different periods. This lack of synchronism in the growth of long bones we have already noted. At 3 years in these babies the distance between knee and wrist is of the order of 21 cm., with extremes at 18.7 and 22.0 cm.

Individual youth.—Figure 69 gives the distance between knee and wrist for seven boys. In all of these curves there is a steady advance, though in most cases with varying slope. The curves advance from around 22 cm. at 6 years to around 35 cm. at 18 years. In some cases a maximum advance is reached at the time of the adolescent spurt of growth or about 11 mm. p. a. The amount of increase in any one year is different in different boys. Thus in the case of no. 3 at 10 years it is 20 mm. in one year, while in the case of no. 7 at the same time it is 8 mm. in one year. The slope tends to become asymptotic during later puberty. In the case of no. 6, the colored boy, change has ceased after 16 years. In the case of nos. 3 and 7 there seems to be a decrease in the interspace between knee and wrist after 16 years, probably due to the fact that the upper limb continues to grow after the lower limb is finished. The greatest interspace is found in no. 5, the acromegalic, indicating that though the radius grows with abnormal rapidity in the acromegalic the femur grows still more

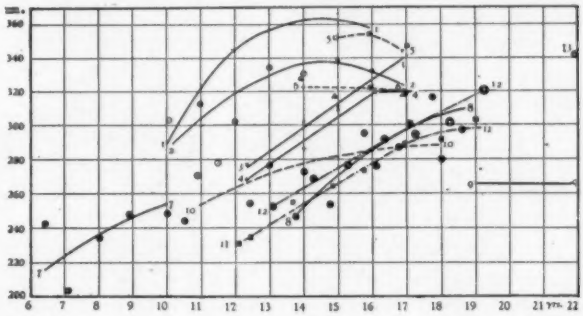


FIG. 70. Individual curves of change with age of interval between stylion and tibiale for twins and some special cases: 1, 2, A. G. and M. G. (monoz. twins), f.; 3, 4, M. K. and G. K. (monoz. twins), m.; 5, 6, E. D. and S. D. (dizygotic twins), f.; 7, A. A., cretin, f.; 8, C. B., treated cretin, f.; 9, E. S., chondro., f.; 10, M. C., atel., f.; 11, A. W., atel., f.; 12, V. R., microcephalic, m.; 13, W. S., chondro., m., aged 29 years.

rapidly; consequently the stylion and tibiale tend to separate (*cf.* figs. 13, 21). The least difference between the two points is found in no. 1, the small Italian boy, and no. 6, the colored boy, whose measurements thus justify the common view concerning the long arms of the Negro.

Twins and special cases.—Figure 70 gives the curves of change in the interval between stylion

and tibiale for two pairs of twins and some special cases. Curves nos. 1 and 2 are of the monozygotic Gar. girl twins; they run nearly parallel, though at a distance apart amounting to 24 mm. after 12 years. Nos. 3 and 4 of the monozygotic boys lie close together being only about 6 mm. apart. Of the cretins curve no. 7 seems to keep a middle position; no. 8, of the treated cretin, lies rather low in the series; no. 9, chondrodystrophic, lies lowest of all and no. 13, of the chondrodystrophic man, occupies at 29 years a median position. Nos. 10 and 11 are of ateliotic dwarfs and these occupy a low position in the series, indicating that in them the arms grow relatively longer than the lower limbs. No. 12 is the curve of the microcephalic boy, V. R. Its position is in the lower quarter of the series, supporting in a way the popular notion of long arms in the case of the low grade feeble-minded. In the case of the Gar. twins, already referred to, the curves slope downward with advancing years after 14. This indicates that in them the arms continue to grow after the femur has become retarded in its growth.

Summary.—The interval between the stylion, arms hanging free, and the tibiale varies in children of 16 years between 26 and 38 centimeters. The interval tends to increase with age from infancy onward, since the length of the trunk plus femur increases faster than does the upper limb. The increase often ceases after the adolescent spurt, but it may continue or there may be a decrease after puberty as trunk and femur grow more slowly while upper limb length continues to increase. The ontogenetic increase of stylion-tibiale interval in children recapitulates phylogeny.

2. PERCENTAGE RATIO OF GLUTEAL DEPTH TO GLUTEAL HEIGHT (GLUTEAL INDEX)

The age change of this ratio has probably not been followed hitherto. The gluteal height is the nearly vertical distance of the posterior-superior pelvic spine from the gluteal fold, as measured by the head-height compass of Matiegka. The *gluteal depth* was measured at the same time that gluteal distance was taken with the same instrument at maximum reading.

The aim of these measurements was to trace the development of the gluteal masses, organs uniquely developed in man because of his erect position. In the anthropoid apes the masses are relatively undeveloped.

General.—The general differentiation curve of this ratio is shown in figure 71. The points taken during the first year are highly variable. However, they indicate clearly that a rapid increase of the index occurs during the first 10 months p. p. from about 23 to 28, or perhaps 30. That is, the depth (thickness) of the gluteal masses is increasing, as it were in anticipation of the upright position. This is another case (of which early childhood has a large number) of early development of an organ in anticipation of its future use (pl. 1, *E*).

After the erect position is assumed the percentage ratio diminishes to a minimum of 22 at 7 years. The mechanism of this loss of 6 points is not clear. Possibly the gluteal fold landmark descends under the influence of gravity and use so that gluteal height is increased; and probably also under the influence of gravity the gluteal depth is reduced as the mass sags. However that may be, the depth increases rapidly and to a great extent between 7 and 11 years when the child's muscular system is developing under the influence of use. In our series after 11 years the depth/length percentage ratio slowly decreases to 24, or less, at maturity.

Sexual.—From figure 71 we might draw the conclusion that during infancy the depth/length index is much less in girls than in boys. This difference is perhaps influenced by sampling, although a difference between means at 9 months of 4 points, where the standard error of

---

PLATE 1

A. Full length portrait of A. deA., m., no. 29, showing slight development of gluteal mass: Index at 16 years, 19. Large stylion-tibiale interval, 349 mm.

B. Chondrodystrophic male, W. S., aged 29 years, showing short humerus and femur (courtesy of Dr. Morris Steggerda).

C. Full length portrait of A. S., m., no. 99, showing protruding development of gluteal mass. Index at 15 years, 25. Also, large stylion-tibiale interval at 13 years, 343 mm.

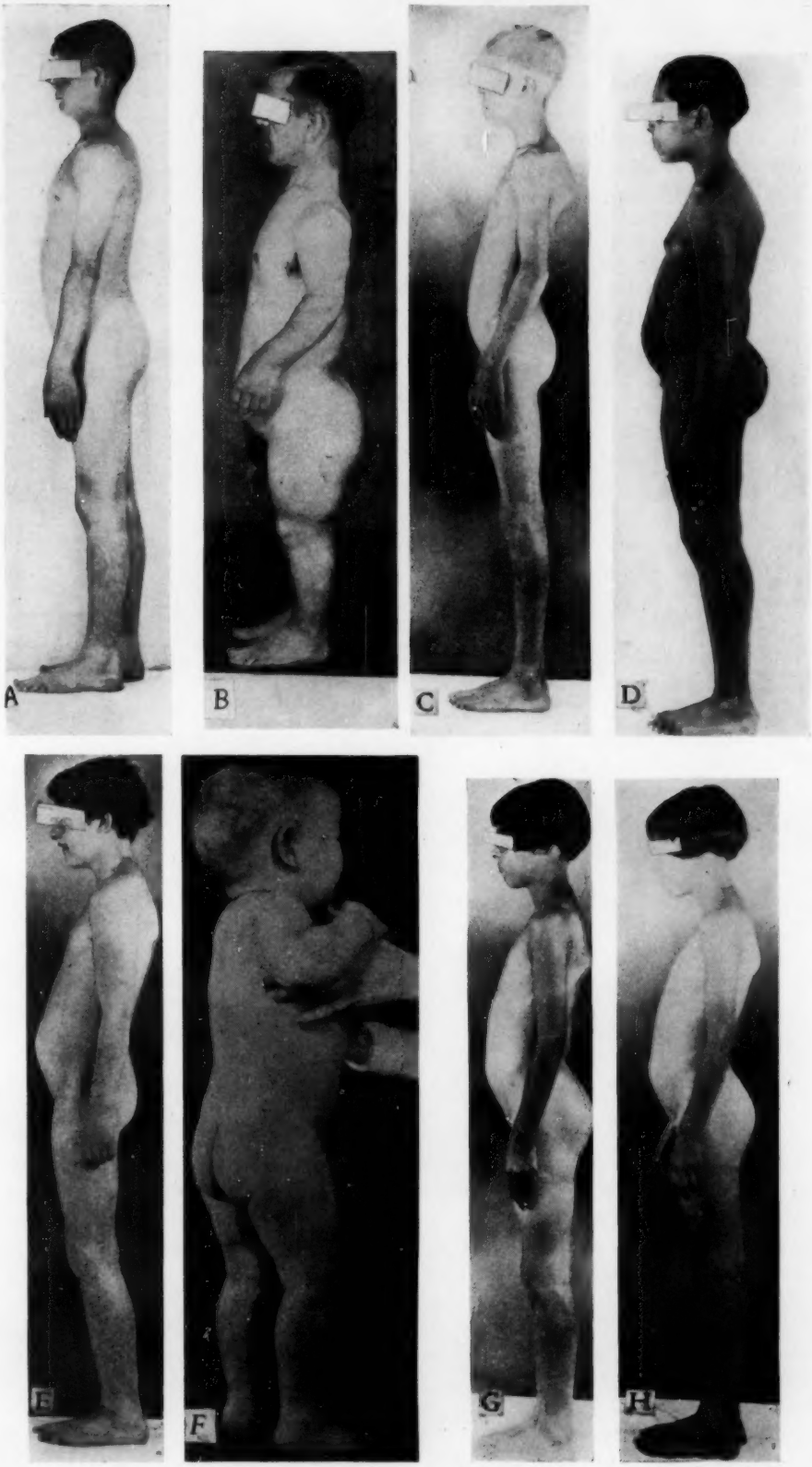
D. Full length portrait of L. D., f., no. 135, showing large development of gluteal mass. Index at 12 years, 27; at 15 years, 34. Small stylion-tibiale interval, at 13 years, 294 mm.

E. Full length portrait of M. C., f., no. 118, showing slight development of gluteal mass. Index at 13 years, 334. Stylion-tibiale interval, 334 mm.

F. Baby showing feeble development of gluteal mass.

G, H. Full length portraits of Mary Gar. (*G*) and her twin sister A. Gar. (*H*) showing similarity of trunk type, upper and lower limbs and their relation to each other. Age 10 years. Stylion-tibiale interval, 303 mm. and 304 mm. respectively.

PLATE 1



the mean is about 1, would seem to be significant. In later stages the curve of girls follows that of boys except that it lies about half a point above until 9 years of age and then falls 2 to 1 point below, with the greatest departure at 11 years. Also, the downward slope of the male index is not found in the female so that a decussation of the curves occurs at the seventeenth year.

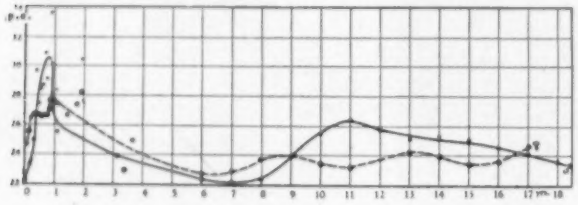


FIG. 71. Standard mass curves of change with age of percentage ratio of gluteal depth to gluteal height.

Racial.—The Negro's index of gluteal depth/length is about 2 or 3 points higher than that of Whites in the case of boys and 2 to 4 points larger in girls (pl. 1, *D*). In the small-bodied Jewish and Italian boys the index is slightly less than in the Nordics.

Summary.—The mean gluteal index increases rapidly in infants until 8 months old, then decreases to late childhood, and thereafter increases slowly to maturity. The small gluteal mass at birth resembles that of young anthropoids. The average boy achieves his maximum index at about 11 years, just before the trunk becomes rapidly elongated at the time of the juvenile and adolescent spurts.

3. GLUTEAL INDEX RELATIVE TO STATURE

Since the gluteal muscles exist to support the erect body, and since the larger the body the greater the need of support, it will be of interest to note how the muscle mass increases with stature.

General.—The data for this ratio are few and scattered. Figure 72 gives the mass curve from birth to maturity excepting between ages 3 and 6 years when the observations are inadequate. In relation to stature the gluteal index (i. e., the percentage ratio of gluteal depth to gluteal height) falls steadily from about 4.6 at birth to 1.9 at 8 years. This means that the gluteal mass develops in thickness less rapidly than total body length up to the beginning of adolescence. This is not difficult to understand inasmuch as the most rapid increase of stature is located in the lower limb bones rather than the head and trunk which is the combination to be balanced. After 9 years, in the males at least, there is an increase in protuberance of the gluteal masses as we saw in figure 71 and this increase is relatively greater in amount than that of stature. This increase seems to be associated with the greater activity of the trunk and upper limb at this age.

Sexual.—During the first 4 or 5 months p. p. the girl babies in my series had this ratio about 0.3 point larger than the boys. Thereafter to 9 years no difference was noted. In girls the index makes no such temporary increase during adolescence as in boys (possibly because of less activity of the trunk); but after 15 years the mean ratio increases in the girls and surpasses that of boys. This increase is associated with, and perhaps due to, the greater accumulation of

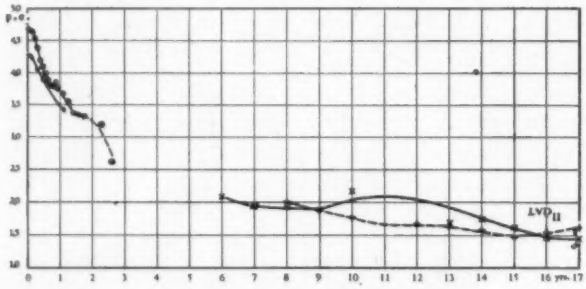


FIG. 72. Standard mass curves of change with age of percentage ratio of gluteal index to stature.

fat on the buttocks in the female. (See discussion in Martin, 1928: 237–241.)

4. GROWTH OF SCAPULA IN LENGTH

The scapular length was measured between the tip of the acromion on the one hand and the end of the spine at the vertebral margin on the other. In the immature, cartilage-bounded, scapula of infants some of these landmarks are imbedded in fat and muscle and are not determinable with great precision but the trend of growth is obvious. The dimension is not standard for bone measurement but seemed the most practicable for measurements on the living.

General.—Figure 73 gives the mass curve of growth from birth to maturity. Beginning at birth at about 59 mm. for boys, length reaches 82 mm. at 1 year p. p. thus advancing 23 mm. p. a.; and 91 mm. at 2 years, thus advancing 9 mm. p. a. in the second year. The advance of the growth curve seems to be slightly interrupted

when the children were first measured standing, doubtless owing to the new position of the bone in the muscles. From 4 to 6 years there are no observations, but growth must be slow. The slope is 4 mm. p. a. at the eighth year, decreases to 2 mm. p. a. at the eleventh year, and then increases rapidly at 8 or 9 mm. p. a. to 155 mm.

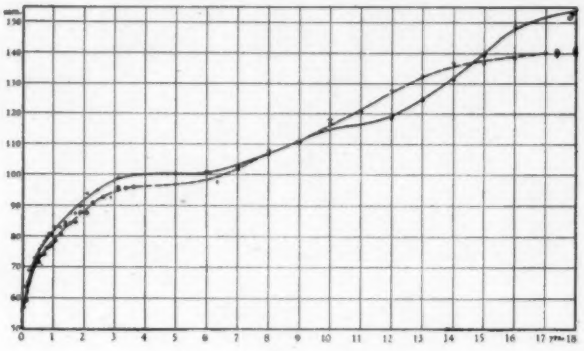


FIG. 73. Standard mass curves of growth of scapula in length.

at 18 years. Increase in length of the scapula keeps pace with the increase in size and activity of the upper limb.

Sexual.—The curve of scapular length of girls lies 2 to 4 mm. below that of boys from birth to 6 years. At 9 years the curves of the sexes decussate and at 12 years the length is nearly 10 mm. greater in the girls than in the boys of my series. However, a second decussation occurs at the fifteenth year, after which the scapula length nearly ceases to increase, standing at 140 mm.; while that of the boys increases steadily to 155 mm. Perhaps this difference is in part due to difference in activity of the upper limb; perhaps in part to difference in the vigor of the growth-promoting hormones.

Racial.—The curve of scapular length of the Negro boys lies always above that of the Nordics —3 to 5 mm. That of the Italians falls below that of Nordics from 16 years onward.

5. CHANGE WITH AGE OF THE ANGLE MADE BY THE PLANE OF THE SCAPULA WITH THE FRONTAL PLANE OF THE BODY

Problem.—In quadrupeds in which the anterior appendages function in the para-sagittal plane the scapula also lies nearly in that plane. Thus in the horse the relations as seen in cross section are nearly as in pl. 2, *B*, 1. In man, on the other hand, the scapula forms a small angle with the frontal plane. In well-developed, upstanding young persons the angle may well be 10° to 0° (pl. 2, *B*, 2). The question is, when does the change of position occur and is it completed at birth?

Only some of the children were measured for this somewhat complicated parameter. But the number was sufficient to give a clear answer to our question.

The method was to measure the distance a between the tips of the two acromial processes and also the distance s between the vertebral margins of the scapulae at the point of intersection with the scapular spine. Then $\frac{1}{2}(a - s)\, l$ = cosine of the angle made by the scapula with the frontal plane; where l is the length of the scapula as defined in section 4. In ungulates the cosine is nearly zero; in the erect human nearly 1.0. In babies of about 3 months p. p. the mean cosine was found to be 0.70; at 16 years it was 0.83 (fig. 74). These cosines correspond to angles of 44° 28′ and 33° 54′ respectively. Thus during the first 16 years p. p. the scapulae on the average pass through 11° of the 90° which transform the infant from a half-way quadruped to a two-thirds or more *Homo*. This is still another example of the

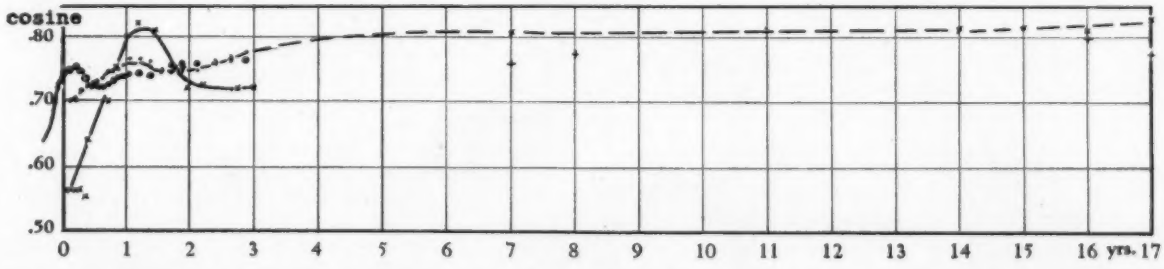


FIG. 74. Standard mass curves of change with age of percentage ratio of cosine of angle made by scapula with frontal plane. The short unbroken curve is an individual one of the baby boy T. P. Prenatal part derived from Embryological Laboratory, Carnegie Institution of Washington.

recapitulation process which speaks indubitably for man's origin from quadrupedal ancestors.

VIII. DISCUSSION

1. THE RELIABILITY OF MEASURES OF LINEAR DIMENSIONS OF LIMBS TAKEN ON THE LIVING

with W. D. Hallock

Every conscientious and critical observer who measures dimensions on the living is impressed by the inevitable inaccuracies, which yield an unknown error to each of his findings. Thus, the value of the numerous curves shown in this paper and of the conclusions drawn from them depends upon the degree of reliability that can be placed on the measurements upon which they are based.

In the case of mass-curves the individual errors, being equally apt to exceed or fall short of the true value, are largely cancelled out and in any case the degree of unreliability can be expressed; but in curves based on "longitudinal" series of measurements of single individuals each error stands unmodified and unknown. The problem is how can we best arrive at the truth of bodily changes from such data?

Four general methods suggest themselves. First, the individual measurements may be repeated in rapid succession as each is being taken. Thus, since stature is so very important for the computation of many ratios, I have been in the habit of taking it three times in rapid succession and using the "best," which is generally the largest. Similarly it would be desirable to repeat a dozen times each measurement on the same individual during one examination. Then the means and their standard errors could be computed and thus the degree of untrustworthiness would become knowable. But this procedure was found to be impracticable because too time-consuming; because the subject, whether baby or older child, could not stand the strain of it; and because the bodily dimensions that were being measured would be altered by fatigue.

The second method consists of the measurement of the same dimension at each examination by two different techniques and making use of the most probable of the values thus obtained. Especially in measuring the length of the extremities this procedure was followed. Thus, with the subject standing erect, arms hanging straight down, the distance of the landmarks from the floor was measured vertically. The length of the segments of the limbs was obtained by difference of the various heights thus found. Immediately thereafter the segments of the upper limb were measured directly between landmarks by means of the long calipers. The latter method gave results which were two to five millimeters longer than those obtained by the former method, for reasons elsewhere stated. The constancy of the difference by the two methods in successive observations gives evidence of the reliability of the individual measurements.

The third method is a comparison of the results of measurements of identical twins. Experience indicates that in such twins there are certain dimensions that grow similarly. One observes how individual measurements differ in the pair. If they coincide or if they lie at a constant distance apart confidence is created in the accuracy of one's measurements. (Compare figs. 3, 7, 12, 15, 18, etc.).

The fourth method is the determination, once for all and for each dimension, of one's error in measuring. Also, one may make some use of

---

PLATE 2

A. Outlines of bones of the left hand of the GAR. twins as seen in roentgenogram, the one in solid line, the other in broken line, showing great similarity in the form-determining factors in twins. The bones are not precisely superimposed in all cases in order that the form of each of the two bones can be distinguished.

B. 1. Diagram of the appendicular bones of a quadruped showing the position of the plane of the scapulae nearly perpendicular to the frontal plane.
2. For comparison diagram of the appendicular skeleton of a human showing the plane of scapulae forming a small angle with the frontal plane.

C. Outlines of the same bone (fourth metacarpal) of a girl B. Mee., with an interval of 6 years between the first roentgenogram (solid line) and the latest roentgenogram (broken line) 1931–37. Slender type of metacarpal.

D. Outlines of fourth metacarpal of a boy, showing profound reconstruction of bone from 1932 (solid line) to 1937 (dotted line).

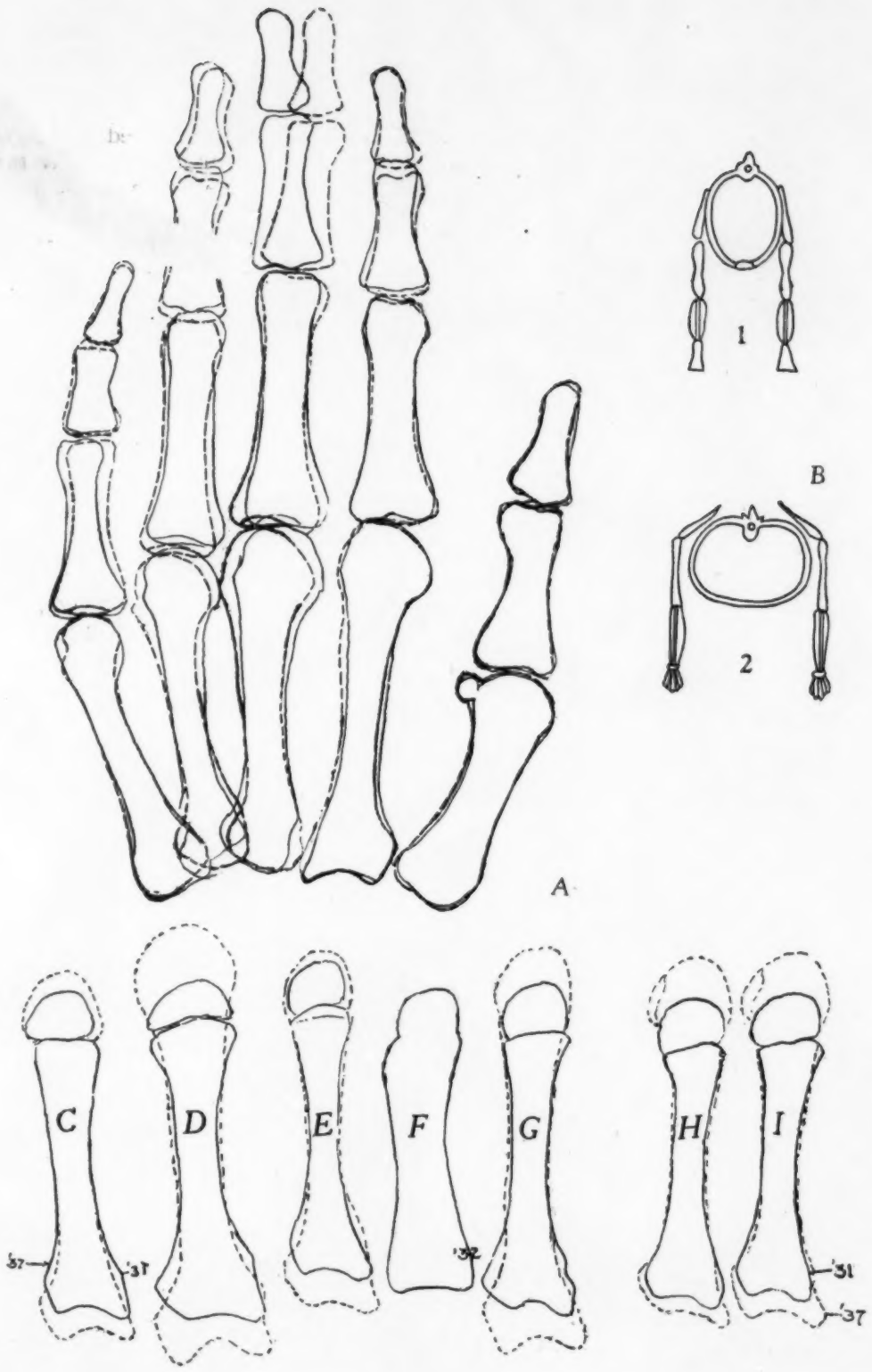
E. Slender type of metacarpal.

F. Gross type of metacarpal of a boy R. Nich. at 13 years of age, showing massive type of bone.

G. Outlines of metacarpal of a child at 1931 and 1937 showing transition of condyle into shaft.

H, I. Second metacarpal bones of GAR. twins. Left, A. G. showing reconstruction of condyles into diaphysis, 1931–1937. Right, M. G. showing similar transformation during same years. The later outline indicated by broken line. R. metacarpal II.

PLATE 2



the error obtained by others in measuring a dimension to determine the intrinsic error in measuring this dimension.

The method of determining personal error is illustrated by a remarkable set of fifty consecutive measurements taken twice daily, alternately at 9 A.M. and 4 P.M. by Dr. Morris Steggerda on one person (R. M. f., 24 years of age) between December 6 and December 17, 1932. The results are given in Davenport, Steggerda and Drager, 1934: 267, 268. For example, the probable errors of the measurement of length of the upper limb and of its segments as taken by Steggerda varied from 1.46 to 5.06 mm.

In indices, Poniatowski (1911) has given the mode of computing the probable error of a value that depends upon other values; and in Martin (1928: 199 etc.) for various ratios is given the "Genauigkeitsfehler" of the absolute difference between two individuals.

Our main problem, in this matter of accuracy, is how significant are the individual values found and how significant are the differences between successive values? In studying the amount of growth of an individual from one observation to the following the behavior of errors is all important.

For this purpose we may make use in the first place of the changes in the successive differences between successive measurements made by Steggerda on the adult mentioned above. In figure 75 are plotted such changes. Thus, if

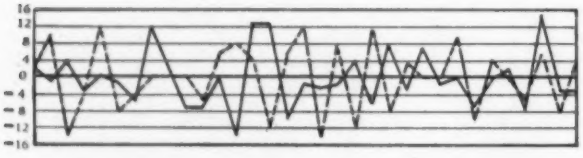


FIG. 75. Illustrating the law of chance errors in measuring on the living body (unbroken line) as compared with chances from throwing coins as described in text (broken line).

the second measurement of a limb bone was 5 millimeters greater than the first and the third measurement was 10 millimeters greater than the second the change in differences between successive measurements was, during this restricted series, +5 mm. If the fourth measurement was 2 mm. less than the third the change in difference was −7 mm. Accordingly the points along the unbroken line in figure 75 give the amount and sense of change in the differences between successive pairs of measurements. The changes run from +15 to −14 mm. The line joining these points may be taken as a random or chance line, indicating that the successive changes are chance changes.

To test this conclusion a series of coin flips was made and the values recorded: the difference between the values of the successive throws was computed and finally the changes in these successive differences were plotted by points on the chart and then the points joined by a dotted line. The coins flipped were: a one-cent piece and three dimes. The sign is determined by the cent piece, heads positive, tails negative. The magnitude is determined by the number of heads appearing in the three dimes cast in the same throw with the one-cent piece, following this scale: 0 head, 0; 1 head, 2; 2 heads, 6; 3 heads, 12. The values are summed and the sums plotted, positive or negative.

It will be observed that in both the lines joining the points, whether determined by measurement or by chance, are zigzag ones, that is, a plus change is usually followed by a minus change either directly or after an interval of one or two throws. Thus in the unbroken line there are out of 32 points: positive changes, 14; negative changes, 15; 0 changes, 3. In the broken line (chance): positive changes, 16; negative changes, 13; 0 changes, 3. The resemblance between the two lines is so close as to justify the conclusion that the changes in sign of the changes in differences in successive measurement of the limb bone length of R. M. are controlled by chance.

But not wholly by a simple chance. By chance, in the pitching of a coin a head will be followed by a head in 50 per cent of the cases; by two heads twice in succession in 25 per cent; by three heads in succession in 12.5 per cent of the cases. That is, we expect in 32 cases a change of sign in successive throws in 16 cases; no change in successive throws in 8 cases; no change in two successive throws in 4 cases. Actually, in the solid line among 32 cases a change of sign in successive cases occurred 22 times; no change in sign 4 times; no change in value 4 times.

Why is this? One reason is that no large "error" is followed by an equal error in the same direction; such would be piling error on error—producing an error of a size such as does not occur. In one case, to be sure, a positive increment of 12 mm. was followed by another increment of 12 mm. larger still, so that the second

"error" was 24 mm. too large. Only some non-recurring accident of posture or other, could yield such a large error. The largest change (not shown in figure) was from the thirty-fifth to the thirty-sixth trial, *viz.*, −24 mm.; partly due to the positive change in the preceding trial. This was immediately followed by a *change* of +15 mm. which brought the measurement close to the true size. A large error in any sense greatly decreases the chance of another change in the same sense immediately following. Such extreme departures from the true value do not occur. For the chances are always best of small successive changes in departure from the true value. After a marked deviation from the true value the chances are for a closer approximation of the true value and this means a change in sign of the zigzag and of a degree that is determined largely by the amount of that marked deviation. Such changes are thus largely compensatory.

The general resemblance of the observed changes in deviations taken on the subject and the changes in successive values based on the flipped coins indicates that the zigzag form of the curves of changes in the "increment" of the successive measurements is so largely determined by chance errors that one may safely draw a curve dividing the distances between successive changes in measurements with some assurance that it is a close approximation to the truth, namely that the body measured is undergoing no real significant change, as indicated by the horizontal zero line.

In cases where the body or its parts is growing the positive accelerations will exceed the negative accelerations and more points will fall above the zero line of no acceleration than below. By superimposing the zigzag curve of acceleration of two organs we may be in a position to determine whether or not acceleration of growth occurs in them simultaneously or otherwise. This matter is further discussed in section 3, below.

Perhaps from the preceding considerations we may draw a conclusion as to the method of drawing the true path of most probable growth in the forest of error. Whenever there are more than two points that deviate similarly and to approximately the same degree above or below the assumed true growth curve, then that growth curve should be altered to pass near such points since these positions are probably not due merely to chance but are significant.

2. GROWTH AND DIFFERENTIATION

In this study postnatal development has been considered from two standpoints: (1) that of absolute increase in size, or growth in the strict anthropometric use of that word, and (2) differentiation, which may be divided into (*a*) relative change in size of organs and (*b*) histological-biochemical changes. In this paper only (1) and (2*a*) are considered.

Absolute growth in the postnatal period is usually represented graphically by an approximately parabolic curve. This curve is subject to many vicissitudes owing to changes in posture (relation of growing part to environment) and to changes in operation of internal growth-directing, growth-stimulating, and growth-inhibiting factors. Such are the curves of growth of the extremities in whole and in part. Examples are shown in figures 1, 4, 8, 13 *et al.*

Differentiation in growth is represented graphically by curves of most varied trend and form. If there is no differentiation in growth between two parts that are being compared, as in a percentage ratio or an index, during a period of time, then the resulting time curve for that period will be a horizontal line. Examples nearly approaching expectation are shown chiefly in mass curves, as in figure 33.

At the other extreme the individual differentiation curves are complex, as for example: figures 40, 41 (forearm to shank length). Especially during infancy individual differential growth is often of bizarre complexity as in figures 40 and 49 (shank 1.:thigh 1.). These complex differentiation curves indicate great independence, specialization, and diversity in development. It is clear that the extreme deviation of these differentiation curves from the horizontal implies the presence of some special disturbing factors. What they are will be clearer after the discussion given in the following section.

Studies such as the present are sometimes called studies in "growth and development." Who invented this phrase I do not know. It is used by T. B. Robinson in his chapter in Abt's *Paediatrics*, 1923. It was adopted as a supertitle for the remarkable series of papers edited by Samuel Brody that has been emanating from the Missouri Agricultural Station since 1926. It received, as it were, official sanction in the White House conference, Part I, Growth and Development of the Child, 1930. In the General

Introduction to the two volumes on this subject published in 1932, the editor (p. 4) says, "We use the two terms, growth and development, advisedly as there is a useful and significant distinction between them. By growth we mean increase in size. As opposed to this, development implies an increase in complexity." This last sentence describes a phenomenon that has been commonly called, for perhaps a century, differentiation. The *Century Dictionary* quotes the *Encyclopedia Britannica:* "Differentiation implies that the simple becomes complex or the complex more complex," thus using almost the terms assigned above to development. Development in biology properly refers to all the processes that lead the egg to become an adult. There may be growth without differentiation and there may be development without growth, but in view of the fact that growth is one of the two most fundamental processes of development, and the universal phenomena of organisms, that by virtue of which all plant life increases and thus permits all animal life to increase, growth may be considered separately, not from development but from differentiation.

That the biochemical processes back of growth and differentiation are quite different has been shown in a series of researches by Gudernatsch and Hoffman from 1912 onward and by Hammett since 1935. Thus Gudernatsch (1934) finds that in tadpoles, growth is promoted by the following amino acids: d-lysine di-HCl, d-arginine and *l*-cystine; while differentiations, on the other hand, are promoted by d, *l*-phenylalanine, *l*-tyrosine, *l*-tryptophane and sometimes by histidine and d, *l*-proline. Hammett, using *Obelia*, found a different set of amino acids from the above both for growth and differentiation. Rose (1937: 298) using rats, reports that essential to growth are: lysine, tryptophane, histidine, phenylalanine, leucine, isoleucine, threonine, metheonine, valine and arginine. Thus while d-lysine promotes growth in vertebrates it is not necessary for differentiation there, and while tyrosine promotes differentiation in tadpoles it is less important for growth. Gudernatsch (1934: 101) states "In the same experiment (*Rana sylvatica*) the three 'growth acids' lysine, cystine and especially arginine delayed more than any other acids the differentiation induced by diiodo-tyrosine, which is again an illustration of the growth *vs.* differentiation antagonism." This matter will come up again in the next section.

The growth of a human child from the egg is a weighty process. During some sixteen years the egg develops into a being of 13,200 times the initial diameter and 30 billion times the initial weight. During some months, growth consists largely of the imbibition of water, so that the fetus comes to consist of 95 per cent water. But it is no longer, as at the start, a sphere. Its axes have differentiated; the poles of its main axis have become different; and from the trunk four extremities have budded out, grown and differentiated. It is the growth of these extremities absolutely and differentially, in postnatal stages, that is the principal subject of this paper.

To consider growth of our series in more detail let us take up that of a long bone; and, first, that of the humerus. As figure 8 shows, during the first postnatal year on the averge the hummerus grows 42 mm. in 365 days. This is at the rate of 115 microns per day. In one baby (J. L.) the humerus grew 45 mm. in 10 months or at the rate of 150 microns per day. Studies made by the Bisgards (1935: 573) show that in goats 82 per cent of the growth of the humerus is at the proximal end. If growth is distributed similarly in the average child the proximal end of the humerus would grow about 95 microns per day on the average or 123 microns per day in the case of J. L. Now this growth occurs at the extreme ends of the bones where the cartilage cells are actively proliferating and laying down, between (or inside) the molecules of collagen and chondromucoid, water, (of which cartilage comprises 55 per cent, according to Vierordt, 1906: 379). This growth of 123 microns per day at the proximal end of the humerus is equal to 14Å (Ångstrom units) per second. According to this computation about one half, or one layer of protein molecules in, or containing, water is formed per second at the growing end of the infantile humerus. However, there is every reason to hold that this growth is not uniform, second after second, but rather that some cells are resting while others are operating much more rapidly than the mean. Just how the cartilage cells build up, with the aid of nuclear enzymes, the collagen molecules out of a few amino acids is a problem for the biochemists to solve.

Differentiation in the long bone is seen in calcification following chondrification, formation of perichondrial cells, and gradual substitution of bone for cartilage, all proceeding *pari passu* with the growth of the bone.

Growth curves like those of figure 9 afford much room for thought. Growth is not proceeding at the same rate in all the babies. No. *H3*, humerus (of a boy), is growing faster than no. *H1* (of a girl). His cartilage cells would appear to be more efficient than hers in forming the intercellular substances in water at the tip of the bone. Also in one and the same baby, growth is not proceeding with the same speed through the years. First there is a slowing up of growth so that in the fourth year on the average only about 3 Ångstrom units are added per second. Secondly, there is a frequent variation, sometimes inside the year, in the speed of growth in the laying down of the formed substance and water of the cartilage. Finally, the bone may even shorten as in girls at puberty (fig. 11)—but that is another story.[5]

The curves of absolute growth give us data for computing the comparative speed of growth for the different bones in infancy and later. These are given in table 9.

TABLE 9

SPEED OF GROWTH OF LONG BONES IN ÅNGSTROMS PER SECOND

Bone	Average					Individual, postnatal			
	Prenatal	1st yr. p. p.	4th yr.	8th yr.	16th yr.	Max. Baby T. P.	Prepubertal Boys	Color. Boy R. H.	Girl A. G.
Humerus	36.8	13.6	3.2	3.2	6.0	20.9	5.1	3.5	4.4
Radius	31.6	12.4	3.5	1.4	3.8	12.3	5.1	7.9	6.3
Femur	58.3[a]	16.8[b]	7.0	5.4	8.6	26.0[c]	9.2	7.0	7.6
Tibia	53.9	15.8	7.0	4.1	1.9	20.9	6.3	5.7	3.5

[a] 504 microns per day; [b] 145 microns per day; [c] 224 microns per day.

The smallest colloidal particles recognizable through the ultra-microscope are about 50 Å. in diameter (Thomas, 1934: 43). Since the increase of cartilage is half water, in prenatal life the femur adds a 50 Å. colloidal layer plus a 50 Å. film of water every second or two depending on whether the water is inside or outside the colloidal particles.

Table 9 shows that, excepting in the colored boy, the femur grows most rapidly of all limb bones at all stages except perhaps the fourth year p. p. Thus on the average the femur grows for six months before birth 58 Å. per second or 504 microns per day; slows up after birth to 17 Å. per second; then to 7 Å. and 5.4 Å. but is accelerated at 16 years to 8.6 Å. per second. In single babies the femur may grow 26 Å. per second and in single boys before or around puberty at the rate of 9.2 Å.

In general the bone that grows fastest will become the longest bone. The quantity as well as the quality of its output is determined both by genes (enzymes) and lyophil substance (protein) and by hormonal stimulants. The femur grows at the fifteenth year at a quite different rate in an acromegalic (8.2 Å. per second) and in a dwarf (5.1 Å. per second) because of constitutional differences.

Curves of gross differentiation (type 2*a*) afford a quite different picture from growth curves. Even mass curves like those of figures 25, 28, 32 *et al* show changes in the comparative amount of growth of the two parts involved. In some cases (fig. 5, curve 6) the relative growth of the dimensions may be fairly uniform for years. In other cases (fig. 30) the relation of the parts is frequently changing. In individuals, these differentiation curves may even assume fantastic forms as in figures 36, 40, 41 *et al.*

In figure 41 the curves show the difference of growth of such apparently unrelated parts as radius and tibia. The relative speed of growth seems in some cases to alternate at intervals of two or three years, in other cases to be for long periods constantly different. Just as the proportion of parts is different in the adult so the parts grow at different speeds before maturity; indeed the latter fact is the cause of the former. Differentiation is an immensely more complicated phenomena than mere growth.

Table 9 shows one other thing—the comparatively rapid growth of the radius in the colored boy. However, the other limb bones of the colored boy do not grow faster than the corresponding bones of the Whites of the same age. The growth of the various limb bones of the one girl considered is not so strikingly different from that of the boys as to warrant any generalization about sex, except that, in general the bones of girls grow more slowly than the corresponding bones of boys.

General bodily differentiation is thus seen to be a very different process from growth. Schmalhausen (1925, 1926) in his remarkable studies *Über Wachstum und Differenzierung* has indeed, like Gudernatsch, found in various animals

[5] Dr. Oscar Riddle tells me that there is increase of blood calcium, probably by absorption from bone, in girls at puberty.

evidence of an opposition between growth and differentiation such as Minot (1913) first pointed out. Schmalhausen has gone into much greater detail than Minot. Thus he concludes (1925: 385):

> Die Wachstumsgeschwindigkeit eines Organs ist durch das Alter des Embryos, durch den Differenzierungsgrad in einer jeden Entwicklungsperiode und durch die zur Zeit auf ihn einwirkenden Substanzen bedingt. Während der Embryonalentwicklung kann man eine periodische Einwirking von Substanzen feststellen, welche das Wachstum des ganzen Embryos wie auch der einzelnen Organe desselben hemmen. Während dieser Perioden einer Wachstumsdepression gehen die wichtigsten Formbildungsprozesse vor sich.

My own studies on the postnatal development of children give some examples that support the conclusion that bodily growth and bodily differentiation tend to alternate. For example in Davenport, Renfroe, and Hallock (1939: 201) the conclusion is drawn that: "The change in rate of basal metabolism probably slightly precedes the spurt of growth in stature." It is as though an increased use of oxygen acts upon the cartilaginous ends of the growing bones so that between forty-five and two hundred days later they are able to show increased growth amounting to perhaps 200 microns within twenty-four hours or, say, 25 Å. per second.

To the problem of how far in general in developing children differentiation and growth are opposed in time my studies do not contribute. But on the problem of partitionment of growth substances or stimuli they do bear as shown in the following section.

3. RHYTHM OF GROWTH

Rhythm in organic processes is a rather trite subject of biological investigation. The word rhythm applies, however, to a great variety of processes. In the sense of Verworn (1922: 459) it refers on the one hand to such changes in the organism as are due to repeated environmental changes (like the growth of rings of trees), to the oscillating potentials following stimulation of the optic nerves, or the regular heart beat. Again, rhythm has been found by students of tumors in the repetition after twenty or thirty years of a tendency to form uncontrolled growths (Hedinger, 1923). This is hardly rhythm since there are only two phases involved.

Rhythm in growth has not been much studied. Stratz (1922: 56) refers to alternating phases of linear and of transverse growth of the body of the child, "Streckung" and "Fülle," of which the latter is repeated three times during life: at infancy, childhood (8 to 10 yrs.), and maturation.

Another kind of rhythm was, so far as I know, first called to attention by Godin, (1903). One of the most important general conclusions drawn by him in his great work on anthropometry of growth is given on page 123:

> La croissance des os longs des membres, procède par périodes alternatives d'activité et de repos qui se succèdent avec régularité. Ces périodes sont contrariées pour les deux segments osseux d'un même membre.

(See also Godin 1935: 189).

It is important to consider the statistical procedure that led Godin to this conclusion. He measured boys at a military school at Saint-Hippolyte des Fort (Gand). There were one hundred and fifty of them, 13 to 14 years of age. First measured in 1896, they were measured at half yearly intervals until 17½ years of age; but some of the boys left school in the middle of this period and new ones were taken in. At the end, only one hundred of the original boys remained. The boys came from different departments of France and her colonies. Thus some Arabs are included. The notations for each individual at each examination numbered 175. The data thus obtained were used for the computation of means. Few data are published on individual boys. Thus it appears that the material is not homogeneous, though the different averages are apparently of the same one hundred subjects.

Now Godin (p. 107) finds that if between two given measurements taken six months apart the femurs of these 100 boys show on the average excessive growth, the tibiae of the same boys within these same time limits show an average retarded growth. The average difference amounts to 8 to 0 mm. or from 2 per cent to zero.

In the case of the two principal long bones of the upper limb the difference in average increment in six months differs from 9 to 2 mm. or 3 to 0.7 per cent. The difference here, as also in the long bones of the lower limb, is most accentuated at 14½ to 15 years, the period when, on the average, the long bones are growing fastest.

Interesting as the conclusions of Godin are it cannot be held that they properly flow from his observations or rather that his method of procedure was capable of yielding a sound basis for his conclusions. Thus in his table of comparative growths of "leg" and thigh (p. 107) he found from $14\frac{1}{2}$ to $15\frac{1}{2}$ years an average increment of the thigh of 8 mm. or a total of 800 mm. in the 100 thigh bones. On the other hand he

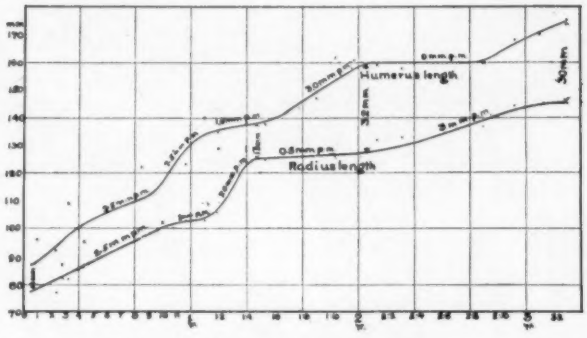


FIG. 76. Growth in length of the radius and humerus of a baby boy T. P. showing alternation of speed of growth in these two bones during the first 3:4 postnatal years.

found during the same interval an average increment of the "leg" of 0 mm. or a total of 0 mm. for the 100 tibiae, since the average is an average of positive values. The sum of 0 mm. growth for 6 months in a total of 100 growing bones is something no one else has found and it can only mean that there were minus values equal to the plus values. But in growing boys of $14\frac{1}{2}$ to 15 years there are actually no minus values to tibial increase. We must conclude that Godin's measurements were taken so crudely that the real positive values were as apt to be recorded negative as positive, and that the law of alternating periods of activity and repose in long bones is not adequately supported by Godin's observations. Difference in growth of bones is an *individual* matter and the final support of Godin's conclusion of rhythm in the growth of the long bones must be sought in measurements made repeatedly on individuals and especially on babies where the bones are growing rapidly and measured at short intervals.

To determine the relation between the growths of the long bones of the same limb in the individual reference is made to figure 76 which gives the growth curves of the humerus and radius of a baby (Thos. P.) during the first 40 postnatal months. With some exceptions the observations were made at first fortnightly, later at monthly intervals. The following points are shown:

First, from birth to 3:2 years the humerus added 88 mm. or 27.5 mm. per year or 8.7 Å. per second. This is a mean increase of slightly less than one protein molecule plus water per second. During this same period the radius added about 65 mm. or 20 mm. per year or on the average 6.3 Å. per second.

Secondly, the rate of increase varies. Thus from 1:2 to 1:4 years the radius grew 16 mm. or at the rate of 96 mm. p. a. or 30 Å. per second. Simultaneously the humerus grew 3 mm. or at the rate of 18 mm. p. a. or 5.4 Å. per second. On the other hand from 1:8 to 2:0 years the radius grew only 1.5 mm. or at the rate of 4.5 mm. p. a. or 1.4 Å. per second, while simultaneously the humerus grew 12 mm. or at the rate of 36 mm. p. a. or 11 Å. per second. That is, the same bone grew at diverse rates especially during the second and third years.

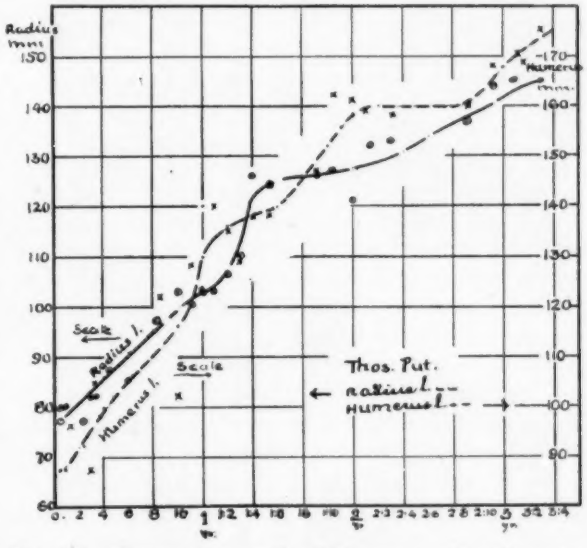


FIG. 77. Curves of growth of humerus and radius of a baby boy T. P. superimposed by changing the ordinate scales showing alternation of speed of growth during the first 3:4 years of postnatal life.

Thirdly, the interval between phases of maximum growth is about 8 months during the second year p. p. and about 15 months during the third year. It seems to be increasing with age.

Fourthly, while one bone is growing faster the other is growing more slowly and this agrees with Godin's conclusions drawn from a quite different and inadequate kind of material.

Figure 77 shows at a reduced time scale and two ordinate scales the same growth curves of the long bones of the upper limb of Thos. P. This figure shows how the lengths of radius and humerus diverge and converge at least twice during the 39 months of the graph.

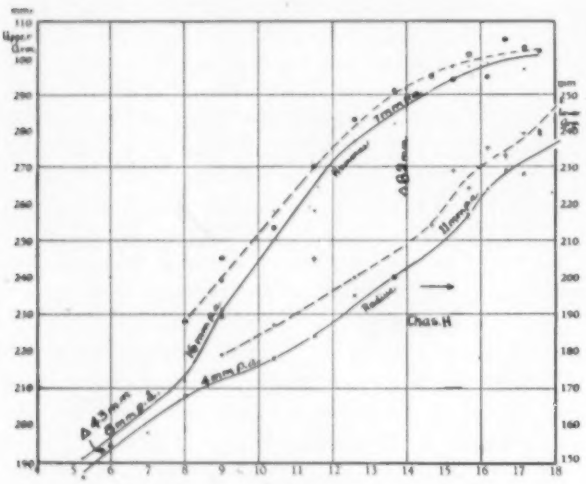


FIG. 78. Growth curves of humerus and radius of a boy C. H., no. 83. Continuous line length determined by subtraction; broken line by direct measurement.

While the foregoing figures illustrate alternations of increment-phases in babies, it remains to consider if this holds for individual children between 8 and 20 years measured semi-annually, embracing the ages covered by Godin's mass series. Figure 78 shows in detail the growth curves of the long bones of the upper limb of such an older boy. Alternating phases are evident here.

If instead of growth curves we consider the changes in increments of the growing bones in order to compare these changes in successive years in the length of humerus and radius we reach a new generalization. Figure 79*c* gives the amount and direction of such changes in one boy. Before 10 years of age the changes in annual increments are in the same sense in both bones—from $6\frac{1}{2}$ to $7\frac{1}{2}$ years both from high to low values and from $7\frac{1}{2}$ to $8\frac{1}{2}$ both from low to high values. But from $8\frac{1}{2}$ to $9\frac{1}{2}$ the change in value of the increment of the humerus is in the direction of decrease; while that of the radius is in the direction of increase. The significance of the lines of this figure will be clear by referring to figure 75 and its accompanying text. But figure 75 refers to errors in the measuring of an unchanging body; figure 79 refers to change in increments of a growing body. Accordingly the changes are mostly positive changes and accordingly mostly above the zero line. By referring to the plus and minus signs in figure 79*a–c* it appears that the plus signs are especially common between $17\frac{1}{2}$ and 21 years in the first, $10\frac{1}{2}$ and $14\frac{1}{2}$ years in the second figure and $6\frac{1}{2}$ and $10\frac{1}{2}$ in the third. In the second figure the changes in annual increment during the period $10\frac{1}{2}$ to $14\frac{1}{2}$ years are strikingly similar in sense and amount.

Figure 80, for five members of the Mea. fraternity, shows a prevalence of alternation of the sense of slope in changes of growth increments.

These examples illustrate the fact that in a long continued individual series the sense and amount of changes of increment in the two bones is sometimes different but sometimes the same—so that Godin's rule does not hold universally. If, now, one examines 247 cases of increments in children and adolescents it appears that, between the ages of 7 and 14 years inclusive, the change of increment slope in humerus and radius is never proportionately more frequently in the same sense than the opposite; whereas, on the other hand, from 15 to 18 years

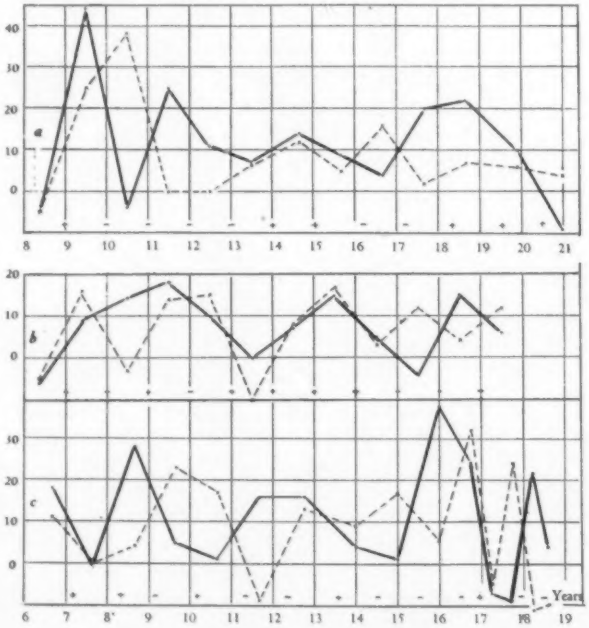


FIG. 79. Curve of change in annual increment at successive ages of "arm" (humerus) length (full line) and forearm (radius) length (broken line) for 3 boys. + positive change; − negative change. Curve *a*, E. H., D. no. 1; Curve *b*, J. G., D. no. 114; Curve *c*, C. D., D. no. 30.

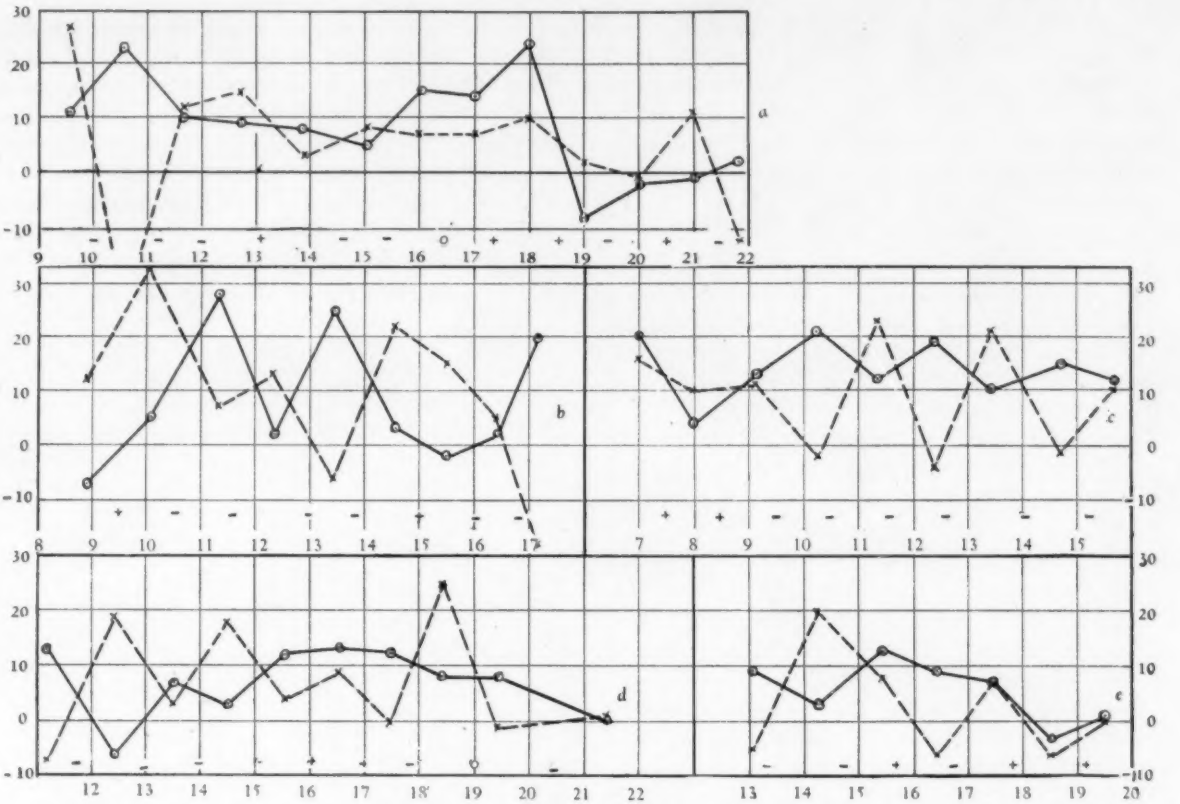


FIG. 80. Curves joining the points of increments in length of the "arm" (continuous line) and forearm (broken line) for 5 members of the MEA. fraternity. Signification of the plus and minus signs as in fig. 79. *a*, G. M., m.; *b*, R. M., f.; *c*, B. M., f.; *d*, W. M., m.; *e*, N. M., f.

inclusive the increment slope is never proportionately less frequently of the same than the opposite. The details are shown in table 10, in which slopes of zero change of increment are omitted.

TABLE 10

PERCENTAGE OF LIKE SLOPE IN SUCCESSIVE CHANGES IN ANNUAL INCREMENTS OF HUMERUS AND RADIUS AT VARIOUS AGES

Age	7	8	9	10	11	12	13	14	15	16	17	18 years
%	44	42	22	22	28	50	39	50	58	60	50	65

From table 10 it appears that during the ages of slow growth of the humerus and radius (9 and 10 years) the speed of changes in growth of the two bones generally alternates, whereas, during the ages of rapid growth (14 to 17 years) the changes in increments are more often in like sense in the two bones.

These facts support the hypothesis that during slow growth there is insufficient colloid material and water for both bones to grow with about the same acceleration; while during the period of rapid bone growth there is, as it were, a sufficiency of growth material and growth stimulus to keep both bones accelerating at the same time.

Another hypothesis, viz., that after the adolescent spurt change in increments in both bones alike are decreasing and so are of the same sense, is not supported by the numerous curves of increment changes that I have made. For often both bones show increase of increments of annual growth even after 16 or 17 years, as seen, e. g., in figure 80, *b*, *c* and *d*.

The conclusion seems justified, accordingly, that alternating rhythm of acceleration holds especially for bone growth during the adolescent stages and is not equally marked at all ages.

That the growth of a long bone proceeds by alternating periods of more rapid and less rapid growth seems supported by the observation that while the growth curve is often wavy the curve of acceleration of increments is almost always a zigzag.

There is another possible hypothesis that must be considered in the alternating slopes of the increment acceleration curves of the long limb bones. This is that among other "errors" to which bone measurement is liable there is that of finding the dividing plane between the two limb-bones. If this is taken too proximal in one measurement it will at the same time result in a too short humerus and too long radius, so that a small increase on the one will be accompanied by a large increase of the other. If at the *next* measurement the land-mark is accurately determined the increase found of the humerus will be too great and that of the radius will be too small. Thus an alternation of sense of increase between the two bones and apparent alternation of speed of growth of either bone will result.

That this is not the whole story is shown by studying the increment-acceleration curves of a lone dimension that has no landmark having a position which may be shared with another dimension. Such is foot length. The individual increment-acceleration curves of foot length (fig. 81) show the same zigzag character as those of the long bones of the limbs. Here surely the changing speed of increments of the curve is due either to changes in the rate of growth, or to accidental errors; possibly both.

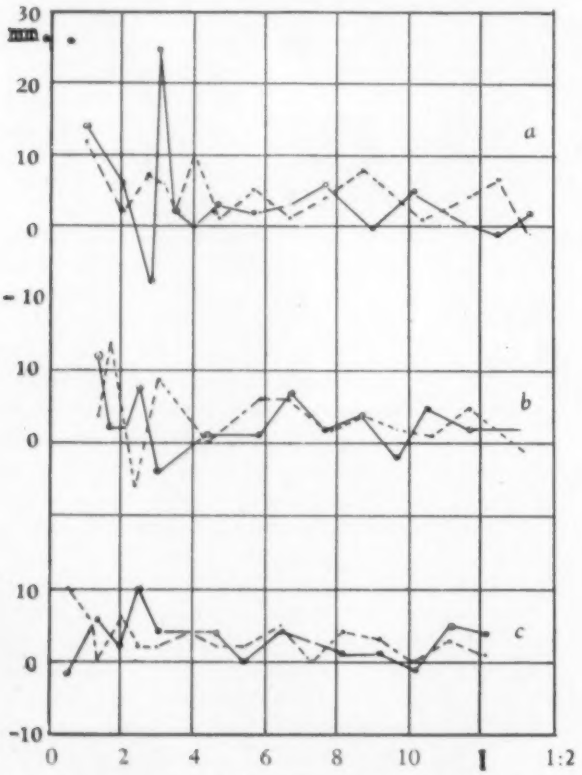


FIG. 81. Curve of monthly change of increments of growth in foot length for 3 pairs of twin babies. Abscissae, ages (years and months); ordinates, millimeters annual change of increment. *a*, McI. twins; *b*, Mil. twins; *c*, Dav. twins; all at Babies Hospital.

The fact that there is a relation between twins (probably monozygotic) in amount and the sense of slope as shown in figure 81, *a*, *b*, *c*, indicates that the growth changes are real and of primary importance. In the case of the twins shown in *a*, the curve made of the broken line anticipates that of the solid line by about one month.

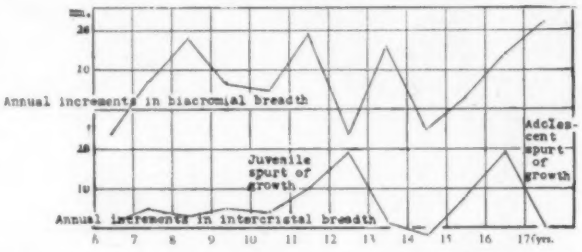


FIG. 82. Amount of change in annual increments in biacromial breadth and intercristal breadth of a boy, J. G., no. 114, showing irregularities chiefly due to juvenile and adolescent spurts of growth.

Another example of rhythmic acceleration is seen in the annual change of increment curves of biacromial breadth or intercristal breadth (fig. 82). The latter shows two maximal changes of increment at the ages of juvenile and adolescent spurts of growth. The former shows two pre-juvenile spurts 3 years apart. There is here an alternation of accelerations not due to an error in locating a dividing plane (as between humerus and radius).

The conclusion that may be drawn from these data would seem to be: adjacent long bones of the limbs do not increase in parallel fashion, at least before adolescence; but a more rapid increase of one is more often than not accompanied by a less rapid increase of the other. Also there are waves of growth increments of the long bones, periods of more rapid growth alternating with periods of less rapid growth, at intervals of 6 to 12 months.

The rhythms of growth of the single long bones depend on rhythmic growth processes at the ends of the diaphysis.[6] That is, there are

[6] For an excellent account of the embryo-histological processes at the end of the growing long bone see Landauer, 1927: 201.

obviously periods of greater and periods of less active formation of colloids and their associated water there. Cell division of the cartilage cells is occurring there, and there is every reason for concluding that during that process metabolism is proceeding slower than in the "resting"—metabolically active—stages. The rhythm in accelerated growth of bones can hardly be due to the rhythm of cell division. For, our rhythms are of a higher order than rhythms of cell activity, since one phase of our rhythm may be of several months duration. The "rhythm" of growth in long bones is more like the variation in bone production with varying health which Harris (1931, fig. 7) has figured. It is probable that all bone growth is rhythmic or periodic or at least intermittent or of rate varying from time to time.

4. THE INFLUENCE OF GRAVITY ON DIMENSIONS AND PROPORTIONS

Inasmuch as the body at all ages is a viscous mass, held together by the skin and slung upon a more rigid bony skeleton, it yields to pressure; and one pressure that is always acting, in the air, is gravity. The body as measured, whether lying horizontal or standing erect, is deformed by gravity.

What is the form of a child? It depends upon whether it is submerged or lying or standing in the air while being measured. In our work the baby was during 11 or 15 months measured lying; then came the time when the child was measured standing. This change in posture frequently caused a sharp break in the curve of growth or of proportions of the parts the growth of which was being followed. This condition is seen in the babies in figure 5 for the upper limb; it also holds for thigh length, while tibia length (fig. 23) shows less effect. Since stature and sitting height are shortened on standing, ratios in which they appear as denominator tend to show a maximum at around 14 to 16 months p. p. Similarly, when the length of lower limb is denominator the ratio between length of upper and lower limb is high at about 14 months (figs. 29 and 40). Other examples will be found in the curves of this abrupt change due to the action of gravity on the bodily dimensions. Of course a correction might be devised to prevent the "break" but it has seemed better to let the readings secured stand for themselves. As a matter of fact the baby who first stands up is shorter than it was a day before when it was recumbent. The form of a mammal is a function of gravity.

5. GROWTH OF LIMBS IN RELATION TO SEX

In general the gross upper limb is longer in boys (except between ages 10 and 15 years) than in girls in accordance with their larger size; but the reverse is true for the lower limb until puberty. At 9 years the average girl is shorter limbed above and longer limbed below than her brother and this excess in the lower limb resides more in the shank than in the thigh (figs. 22 and 20).

While in relation to stature the gross upper limb is much the longer in boys (fig. 25) in relation to sitting height this difference largely disappears in girls because of their longer trunk. On the other hand the gross lower limb in relation to stature is greater in girls than in boys, while in relation to sitting suprasternal height it is still greater to 10 years of age, but thereafter less because of the long trunk of post-adolescent girls.

In relation to net lower limb length the net upper limb is much (3 per cent) greater in boys than in girls. The "arm" forms practically the the same proportion of the upper limb in both sexes; while thigh length to net lower limb is slightly the greater in the male; and tibia is proportionally the greater in the female.

In relation to stature, span is about 1 per cent greater in the male. Girths of the "arm" and of the calf are greater in the mean girl than boy and that of the thigh is absolutely 5 to 12 per cent greater in the girl and 7 to 12 per cent greater in relation to stature. In general the ankle girth is smaller in girls than boys.

Hand length is about the same in girls and boys but width is greater in the latter; consequently the hand index is greater in boys.

The mean distance from stylion to tibiale is about the same in the two sexes. The gluteal depth in relation to height is greater in girls than boys except between ages 9 and 17 years.

Thus, using the male average boy as standard, girls have shorter upper and longer lower limbs and especially longer tibiae; they have longer lower limbs in relation to stature; the thigh forms a smaller fraction of the limb, and the tibia, conversely, a larger. Girls, because slender boned, have relatively slenderer hands and ankles.

6. GROWTH OF LIMBS IN RELATION TO SOCIAL TYPE

The mean lengths of the gross and the net upper limb (humerus and radius) increase, with general mental and physical development, in the order Mongoloid, idiot, feeble-minded, standard. The lengths of the lower limb also decrease with decreasing intelligence. The thigh length is greatest in the control series. The thigh of the Mongoloid at 7 years is only 1 cm. shorter than control, grows much more slowly than the control from 7 to 12 years of age, then grows almost as rapidly as the control until growth stops at 17 years. The tibia of the Mongoloid likewise departs with age from the control.

In the matter of percentage indices: intermembral index of the Mongoloid group is slightly higher and, after adolescence, that of idiots is definitely higher than the control. The relative upper limb length is greatest in the control series; that of the Mongoloid dwarfs slopes downward from the control with age. The length of net upper limb in relation to length of net lower limb before 16 years of age is largest in the control series, but with greater growth of the lower limb in the control series the index becomes thereafter relatively low. The Mongoloid dwarfs have a medium index because of short upper limbs; the idiot group a large index because of long upper limbs.

"Arm" length in relation to gross upper limb length is low in the control series; but the Mongoloid dwarfs have a still lower ratio. In the radio-tibial index the controls, idiots, and Mongoloids, form a decreasing series. The relation of thigh length to the rest of lower limb length was highest in the D2 series; lowest in the Mongoloid dwarfs. The crural index is low in the control and high in the Mongoloid dwarf series because of relatively medium tibias and short femurs in the latter. Tibia length in relation to stature is least for the Mongoloid dwarfs, while the controls occupy a high place. Span in relation to stature follows the descending series: control, D1, idiots, and Mongoloid dwarfs. Calf girth in relation to stature after adolescence is greatest in the control group and fairly high in the Mongoloids, because of dwarfism.

As for the hand, its index is small in the control series and large in the Mongoloids. The hand area is largest in the control, smallest in the Mongoloid series.

Thus the control boys rank highest in limb lengths and are low in intermembral index, have greatest relative upper limb index and upper to lower limb length until surpassed by the idiots after adolescence; have the "arm" forming smaller portion of upper limb than in most other series; have a low crural index and high relative span. The hand is slender.

7. GROWTH OF LIMBS IN RELATION TO RACE

As is well known the Negro race has longer limbs than the White. The present study analyzes the matter further, using Negro-White hybrids. Comparisons relative to stature need not be considered since the mean stature of White and Negro children is about the same. White is here taken as standard for reference, i. e., control.

The "Negro" children have upper limbs 2 to 6 cm. longer than Whites or about 8 to 10 per cent. Humerus at 14 years is 1.8 cm. longer or 6.4 per cent. The radius is about 1.8 cm. longer or 8.6 per cent; i. e., the excessive length of upper limb of the Negro lies chiefly in the radius.

The mean lower limb of the Negro is about 5 cm. or nearly 6 per cent longer; the excess is greater in the upper than in the lower limb. The mean thigh (femur) is about 2.4 cm. or 6.3 per cent longer. The tibia is about 2.6 cm. in excess, or 8 per cent. Thus the radius and tibia are the limb bones the length of which differentiates most the Negro from the White. The order of excess is about as follows (highest first): radius 8.6 per cent, tibia 8.0 per cent, humerus 6.4 per cent, femur 6.3 per cent.

The intermembral index is a point and a half greater in Negroes than Whites, owing to the long radius and hand of the former. The net intermembral indices of the two races are very similar.

The femoro-humeral index is about 2 points greater in Negroes than in Whites. The brachial index is 7 points higher in Negroes.

The relative lower limb-length is 10 points higher in Negroes.

The thigh length in relation to net lower limb length is 1.7 points less than in White—the Negro has an especially long tibia.

The crural index of Negro boys is high, on account of long tibia.

The calf girth of the Negroes is about 1 cm. in excess, but the minimum girth is no larger.

The thigh girth in my group was 4 to 8 cm. in excess of that of Whites.

The mean hand is 0.5 cm. longer in Negroes, and narrower, so that the hand index is low.

Thus in relation to the Whites the Negro has, first, long upper limbs and, secondly, long lower limbs. The distal segments are especially long, as in the gibbon.

8. INDIVIDUAL LIMB GROWTH AND PROPORTIONAL CHANGE

Babies.—The individual growth curves of babies are of diverse types.

(*a*) The simplest is the paraboloid growth curve from birth to maturity, e. g., the curve of growth of the foot length (fig. 64), with minor fluctuation.

(*b*) A complication is introduced by gravity as in the curve of growth of thigh length.

(*c*) The curve of extremely rapid growth-fluctuation in speed is seen in the thigh-girth.

(*d*) Wavy growth of bones of upper limb, reciprocating, alternating growth (fig. 9).

(*e*) The curve of slight differentiation is shown in the percentage value of gross upper to lower limb (fig. 29). Except for the break due to gravity the curve is relatively smooth, nearly level, with a slight slope down with age.

(*f*) Extremely different is the differential curve of forearm in relation to "arm" (fig. 36). Here all curves are thrown into a series of waves, which condition agrees with what we have earlier seen to be a consequence of the rhythm of growth. When the radius is growing rapidly the humerus is growing slowly, and vice versa; so that a great exaggeration of variations in growth of the single bone occurs.

It is obvious on inspection of these figures (paragraphs (*a*) to (*d*)) that there is no universal "law of growth" in babies' limbs. Each part of the child grows in its own fashion and is influenced in its growth by that of neighboring parts.

Individual youth.—Figure 83 shows five graphs revealing different types of curves found in growth and differentiation.

(*a*) Curves of absolute increase in length of the shank in boys between 5 and 21 years. On the whole the advance is steadily upward but not strictly linear in any case. There is a slight waviness indicative of varying speeds of growth. The growth rate is high at the fifteenth and sixteenth years (period of adolescent spurt) and tends to be low at twelfth and thirteenth years (from fig. 24).

(*b*) Growth curves of a small dimension—height of internal malleolus. The curves approach linearity, but with some breaks. There are no obvious waves.

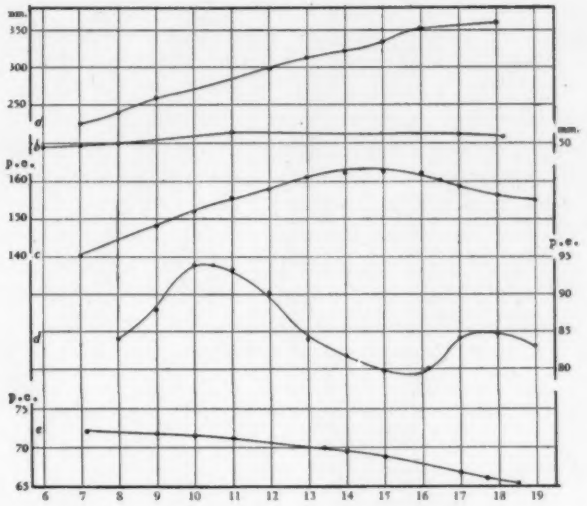


FIG. 83. Five graphs revealing different types of curves found in growth and differentiation.

a. Illustrates the curves of fairly uniform absolute increase but with a fine wave of varying of growth: length of the shank of a boy between 7 and 18 years. Fig. 24 (1).

b. Illustrates the growth curves of a *slowly growing* dimension: height of inner malleolus.

c. Illustrates the *convex above* type of curves of change with age of the percentage ratio of gross upper limb length to sitting suprasternal height for boys. Convexity due to greater growth of the limb than of the trunk until after the adolescent spurt.

d. Illustrates *wavy* type of change with age of percentage ratio of shank length to thigh length due to reciprocal rapid and slow growth of the two bones. Fig. 50 (4).

e. Illustrates curve of slowly decreasing percentage ratio as seen in minimum girth of forearm to maximum girth of forearm.

(*c*) Curves of percentage ratio of gross arm length to trunk. These curves show such irregularities as to suggest variations in speed of growth of arm in relation to stature. Some of the great convexity is due to large adolescent spurts of limb length.

(*d*) This ratio between shank length and thigh (radius:femur) shows some striking differentiation with advancing age. The slope is in general downward, since the thigh grows faster than the shank (fig. 50, no. 4).

(*e*) Change with age of minimum girth of forearm in relation to maximum girth of forearm shows smooth curves sloping strongly downward.

Figure 83 justifies the conclusion that, even after childhood, growth is occurring at various rates in different parts and hence producing adult organs of different size and shape.

9. A COMPARATIVE STUDY OF TWINS

Twins, as I studied them, whether as babies or adolescents, were already the end products of some months (in one case) or years (in the others) of development. In my measurement of the child I was determining (within limits imposed by the methods) the size and proportions attained after this long period of development from the apparently simple egg. This gives some sort of a measure of the efficiency of the gene-cytoplasmic reaction in controlling the development of a child.

Twins, as everyone knows, are either dizygotic ("fraternal") or monozygotic ("identical"). These studies made it possible to contrast the performance of dizygotics (with unlike heredity but similar environment) with that of monozygotics (with like heredity and also like environment).

Among the six to ten babies whose course of development was outlined was one pair of twins: Florence and Margery McI. The results of measuring one ratio are shown graphically in figure 54. A careful comparison of the determinations of length of thigh and girth of ankle, and of ratios like upper limb length to sitting suprasternal height, and thigh girth to stature impresses one with the fact that the developmental details are synchronized strikingly to produce so nearly identical courses of growth and differentiation. Take, for example, growth in thigh length. This dimension is not certainly determinable in active babies closer than 10 mm.; yet with two or three exceptions

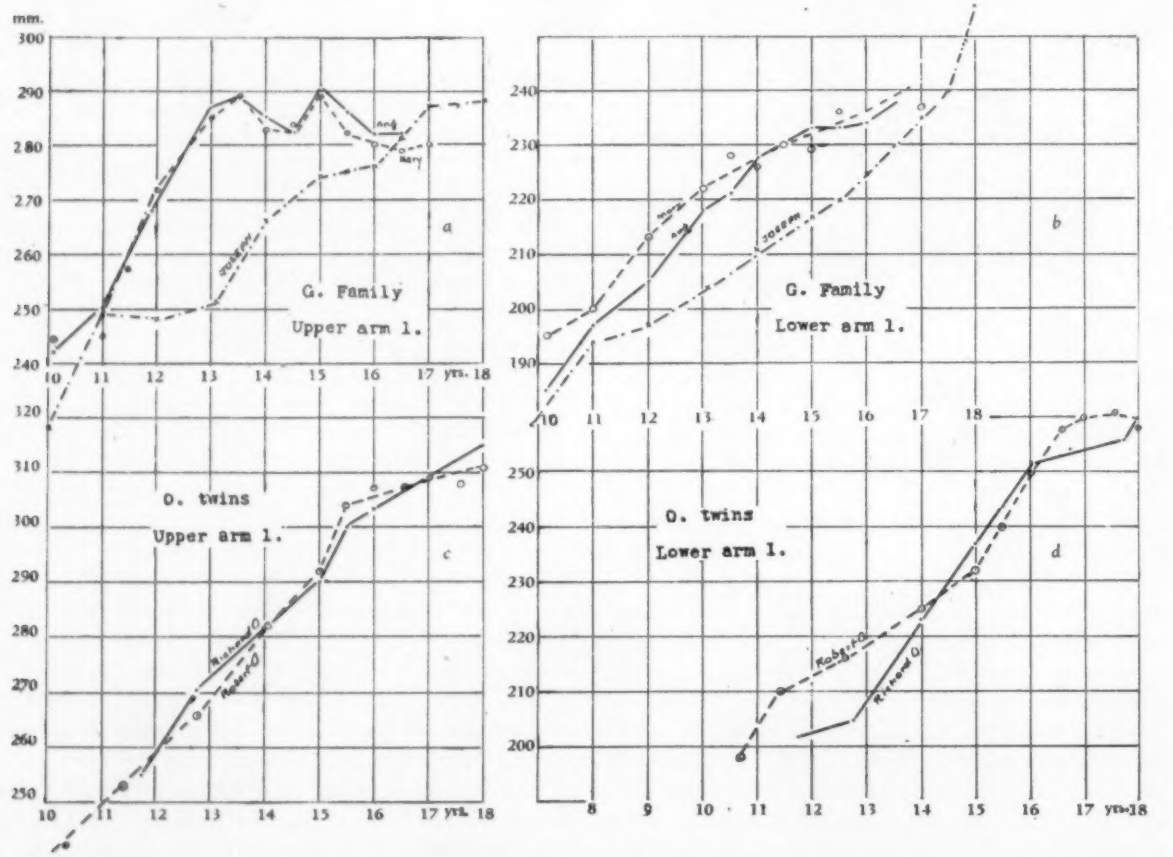


FIG. 84. Comparative growth of "arm" and forearm in twins. In the case of the G. fraternity the curve of the growth of a brother is also shown as a control.

the length attained (as measured) during the first 3 years p. p. agreed within that limit. Even more striking is it to have a ratio as between thigh girth and stature closely similar for the twins at all ages from birth to 3 years. At the end of nine months of intrauterine development the percentage ratios of thigh girth to stature of these babies depart less than 1 per cent from identity, and although these babies had some marked illnesses the above ratio from 4 months p. p. to 38 months decreased only from 37 to 33 per cent. In some cases the attempt to superimpose the curves of the twins was abandoned as impracticable because findings through the months so often coincided. On the other hand, in a few ratios the monozygotics were not constantly the same. The internal control in these cases is inadequate to produce identical results.

The other twins measured were the monozygotics: Angelina and Mary G.; George and Michael K.; Richard and Robert O.; and Daniel and Ralph S.; and the dizygotic colored girls, Evelyn and Sarah D. By every test of general appearance, finger prints and dermatoglyphs the decision as to monozygosis was supported. Curves of some or all of these twins are shown in figures 3, 7, 12, 15, 18, 27, 44, 51, 56, 58, 60, 62, and 66. As one looks at these figures parallel lines strike the eye, and these will generally indicate the course of growth or differentiation of monozygotic twins. Examples are seen in figure 12, nos. 1, 2; figure 15, nos. 7, 8; figure 18, nos. 1, 2; 3, 4; 7, 8; figure 84.

Striking are the differences between the curves of fraternal twins. Examples are seen in figure 3 (3, 4); figure 7 (3, 4); etc. Generally the curves of these fraternal twins run only roughly parallel, as curves of sibs do.

In some dimensions or ratios the curves even of monozygotic twins do not have a similar course. Some traits are less completely under genetic control than others. These facts are brought in detail in table 11.

In general, in the simple dimensions, the monozygotic twins grow in nearly the same way. Exceptions are the thigh and shank lengths (which are influenced by gravity) and hand length. The curves of girth, especially, run parallel. Among the ratios, those involving gross upper limb length appear more frequently in the non-parallel than in the parallel curves.

TABLE 11

CURVES THAT ARE ALIKE AND THOSE THAT ARE UNLIKE IN MONOZYGOTIC TWINS

Curves close or parallel	*Curves distant or non-parallel*
Gross upper limb length	Thigh length
Net upper limb length	Shank length
"Arm" length	Gross upper limb length / Stature
Forearm length	Gross upper limb length / Sitting suprasternal height
Gross lower limb length	Gross upper limb length / Gross lower limb length
Gross upper limb length / Sitting height	Net upper limb length / Net lower limb length
Net lower limb length / Sitting height	"Arm" length / Gross upper limb length
Forearm length / "Arm" length	"Arm" length / Thigh length
Gross lower limb length / Sitting suprasternal height	Forearm length / Shank length
Shank length / Gross lower limb length	Shank length / Gross lower limb length
Net shank length / Thigh length	Thigh girth / Stature
Span / Stature	Hand length
Thigh girth	Hand width
Girth of right calf	Hand length / Hand length
Girth of right ankle	Gross upper limb length / Stylion minus tibiale height
Girth of "arm" / Gross upper limb length	Stature
Min. forearm girth / Max. forearm girth	
Thigh girth / Gross lower limb length	
Hand width	
Hand area	
Hand length / Forearm length	
Stylion minus tibiale height	
Foot length	
Foot width	

10. SPECIAL CASES THAT REVEAL MORPHOLOGICAL ABNORMALITIES

A. Dwarfism

In any institution for the feeble-minded a great variety of undersized persons will be found. Indeed the great majority are below the average in all dimensions, as a comparison of the control with the LVD series, shown on many of the figures, reveals.

The principal types of dwarfs are chondrodystrophic, Mongoloid, hydrocephalic, cretinous, rachitic, and others—the so-called ateliotic,

infantile, and microsomic dwarfs. These last are largely overlapping types, or in some cases the terms are synonyms. At any rate these last three are poorly defined. They have in common small stature, usually under 125 cm. for females, 130 cm. for males. A good discussion of the problem of classification of dwarfs is to be found in O. Maas, 1920, 1923, 1925. Maas reaches the general conclusion that dwarfism is rarely due exclusively to defects in hormones from thyroid, anterior pituitary or parathyroid; or to deficiency of Vitamin D. Rather, if an endocrine unbalance occurs, two or more endocrine organs may be involved so that the end result is due to several causes. It may be added that extreme undernourishment during childhood or a defect in the gene-cytoplasmic interchange may also be a cause of inadequate growth.

The curves of special cases are usually graphed in the same figures as those of the twins:

(*a*) *Cretins.*—Of these, five were examined and three of them followed during 5 to 2 years. The latter are Arline A., b. Jan. 1928; Howard L., m., b. Oct. 1931, and Claribel B., f., b. 1918. Of these Arline and Howard had received no thyroid extract; while Claribel had received some intermittently. At 9 years Arlene's stature was 13 cm. less than standard for that age while Claribel's at 13: 10 years was 20 cm. less and at 17:9 years 13 cm. less. Howard at 5 years was only 3 cm. below standard. He is stated to have had a Korean father and White mother.

From table 12 it appears that in lengths both cretins are below standard and usually the untreated cretin (Arline) departs further from standard than the treated one. But in hand width and in girths the reverse is true. The greatest departure from standard is seen in the short lower limb and especially the small thigh girth. Since the upper limb is less shortened than the thigh the distance from stylion to tibiale is reduced. It is noteworthy that the length of thigh is less affected by therapy than the length of the "arm." The small limb girths of Claribel are perhaps partly a personal idiosyncracy, and partly due to her longer limbs.

(*b*) *Chondrodystrophics.*—Chondrodystrophy is a condition of dwarfness that is due to malformation of the limbs ascribed to retarded or abnormal development of the cartilage-precursor of the long bones. The cause of chondrodystrophy (or achondroplasia) is unknown. The explanation of Jensen (1920) though plausible remains unproved experimentally. Among the best studies that we have on chondrodystrophy are those of Landauer (1927) and Dunn (1927). The conclusion reached by them is that it is most

TABLE 12

THE DEPARTURE OF ARLINE AND CLARIBEL AT 10 AND 14 YEARS RESPECTIVELY FROM A CONTROL (ANGELINA G.) FOR VARIOUS DIMENSIONS AND RATIOS

Dimension	Arline	Claribel	Ratio	Arline	Claribel
				Percentage ratio	
Gross upper limb length	−10.5 cm.	−7.5 cm.	Gross upper limb length/sitting height	−9	6.5
Net upper limb length	−8.7	−7.6	Gross upper limb length/sitting suprasternal height	−17	—
"Arm" length	−5.1	−4.5	Gross upper limb length/gross lower limb length	3.7	1.2
Gross lower limb length	−15.0	−13.0	Net upper limb length/sitting height	−7.2	4.0
Thigh length	−8.0	−7.6	"Arm" length/thigh length	5.5	0.6
Shank length	−8.0	−5.1	Forearm length/"arm" length	4.5	3.3
Thigh girth	−0.4	−14.2	Gross lower limb length/sitting suprasternal height	−22	9
Girth, right calf	−0.5	−7.2	Net shank length/gross lower limb length	−1.4	1.0
Girth, right ankle	−0.4	−5.0	Net shank length/thigh length	−2.5	3.5
Right hand length	−2.1	−1.7	"Arm" girth/gross upper limb length	10	−3.2
Right hand width	−0.3	−0.8	Minimum girth, forearm/maximum girth, forearm	1.2	2.1
Right hand area	−17 cm.²	−18 cm.²	Thigh girth/stature	4.4	−4.5
Stylion minus tibiale height	−3.4 cm.	−11.0 cm.	Thigh girth/gross lower limb length	11.5	−11.4
Foot length	−2.5	3.7	Hand width/hand length	4.4	0.5
Foot width	0	−1.1	Hand length/gross upper limb length	1.4	1.2
			Hand length/forearm length	3.6	3.8
			Stylion minus tibiale height/stature	2	−4.0

probably due to "a deviation from the normal chemical constitution of the egg which is maternally determined in response to some periodically variable agent." In the case of mammalian chondrodystrophy we should have to think of an abnormality of the internal secretions passing through the placenta. In any case, growth of the long bones is due to formation of more collagen and water at the cartilaginous growing ends of the bones when appropriate amino acids are provided. This growth is accelerated by hormones apparently both from the thyroid in childhood and from the pituitary at adolescence. In chondrodystrophy this terminal growth is subnormal.

At Letchworth Village seven cases of chondrodystrophy [7] were examined. The most striking case was that of a man (W. S.) about 30 years old. There were also twin sisters, Elizabeth and Katherine S.

Taking Walter S. (a rather extreme type) at 29 years (pl. 1, *B*) and comparing him with standard boys of age 20 the differences shown in table 13 (in cm. or per cent) appear.

The principal defects in development of this chondrodystrophic individual are in the lower limb more than the upper limb, absolutely but not relatively. The span/stature index is 10 points below control and the net upper limb length/sitting height percentage ratio is 33 below control. On the other hand the hand length/forearm length is a high percentage ratio, viz. 27 points above control. In a word, it is the long bones that are most affected in their linear growth.

(*c*) *Mongoloid dwarfs.*—These striking and much studied dwarfs constitute about 5 per cent of an institutional population (Tredgold, 1915: 212). Their etiology is very uncertain. Such hypotheses have been suggested as: fetal children, products of the exhausted uterus (Shuttleworth, 1909); cretinoids (Mitchell, 1876); thyroprivias (T. Smith, 1896); parasyphilitics (Sutherland, 1899).

Mongoloids are often the only affected member of a large fraternity; the incidence of Mongolism in fraternities increases with the age of the mother. Twin Mongoloid dwarfs are rare. However, since twins occur in only about 1 per cent of the population of an institution and since Mongoloids constitute only about 5 per cent of the institutional population, this rarity is not surprising. By singeing the thyroids of pregnant mothers Swingle and I were able (1927) to obtain litters of dwarf mice. As a result of studies made by Grace Allen and me (1925) the provisional conclusion was reached that a Mongoloid dwarf is a child who did not receive the essential developmental impulses (hormonal catalysts) during fetal life and is not provided with an endocrine mechanism for making good entirely these deficiencies in later life.

From table 14 it appears that the gross lower limb is slightly more shortened in these dwarfs than the gross upper limb, but the relation of net upper limb length to net lower limb length is practically the same as in the control. Simi-

[7] The question of diagnosis arises. At Letchworth Village there are not a few cases of curved bones of the lower limb indicative of rickets. But the cases of short upper and lower limbs, knobby, especially at the joints and with proximal segment especially shortened, seemed an easily distinguishable if not easily definable group. The shallow root of nose was another characteristic. See O. Maas, 1920.

TABLE 13

DEPARTURE OF DIMENSIONS AND RATIOS OF CHONDRODYSTROPHIC MAN FROM CONTROL

Gross lower limb length	−51.0 cm.	Net upper limb length/sitting height	−33.4%
Gross upper limb length	−35.5	Gross upper limb length/gross lower limb length	−10.5
Thigh length	−28.0	Span/stature	−10.4
Net upper limb length	−26.0	Gross upper limb length/stature	−6.5
Net "leg" length	−22.6	Gross upper limb length/sitting height	−6.0
Humerus length	−15.9	"Arm" length/gross upper limb length	−3.1
Calf girth	−9.3	Net upper limb length/net lower limb length	0.0
Ankle girth	−7.6	"Arm" length/thigh length	+3.0
Foot length	−7.0	Stylion minus tibiale height/stature	+4.0
Thigh girth	−6.8	Hand width/hand length	+4.3
Hand length	−5.4	Hand length/gross upper limb length	+9.0
Foot width	−3.0	Thigh girth/stature	+12.6
Hand width	−0.6	Thigh girth/gross lower limb length	+24.0
Stylion minus tibiale height	+5.0	Hand length/forearm length	+27.0

TABLE 14

DEVIATION OF MEAN DIMENSIONS OF MONGOLOID DWARFS FROM CONTROL, AT 16 YEARS OF AGE

Stature	−22	cm.	
Gross upper limb length	−11		
Humerus length	−8.4		
Radius length	−4.4		
Gross lower limb length	−13.3		
Thigh length	−8.3		
Shank length	−4.9		
Hand length	−3.5		
Hand width	−3		
Hand area	−38	cm.²	
Foot length	−4.3	cm.	
Foot width	+2.0		
Gross upper limb length / Sitting height	−2.6%		
Upper limb length / Lower limb length	−8.0		
Net upper limb length / Net lower limb length	−0.2		
Humerus ("arm") length / Gross upper limb length	+0.7		
Radius length / Humerus length	−2.5		
Thigh length / Net lower limb length	+3.5		
Shank length / Thigh length	+2.9		
Hand index	+2.5		
Foot index	+7.5		

larly the thigh length is (very slightly) more shortened than the humerus and the shank length than the radius The hand and foot of the Mongoloid are each about 4 cm. short, but the foot is absolutely slightly wider. The index (width/length) of the hand of the Mongoloid is greater than that of the control; and much greater in the case of the foot. The reduction in length of the long bones is much slighter in Mongoloid dwarfs than in chondrodystrophic dwarfs.

(*d*) *Other dwarfs.*—The principal other dwarfs approximate the microsomatic (or "ateliotic") type. Of these the growth was followed for the longest time of the following: Marion C., f., 11–19 years, U. S.; Lillie S., f., 10–18 years, Jewess; Anna W. and Mary W., her sister, Irish extraction, 12 to 22 years. At 14 years, Anna was 121.4 cm. tall and Mary 115.5 cm. tall (standard 152 cm.). Since Anna's series is the nearest complete, her growth curves are usually plotted in the figures of this paper. It may be added that besides these sisters their fraternity included a sister born badly crippled and deformed, who died at 2 years in convulsions, two brothers undersized and microcephalic. The mother is believed to have had numerous miscarriages, had a brother who died insane, and other relatives who were erratic.

These sisters did not seem to be of the marked infantile type, since secondary sex characters appeared, though retarded. Thus Anna W. at 13 years showed no breast development but this began at about 14:6 and was marked at 15:2 years; pubic hair appeared at 16 years. Mary's breasts developed at 14:6 years and pubic hair at 16:8 years. Lillie S. at 14 years was 120 cm. tall; breasts developed at 15 years. Marion C. at 14 years was 125 cm. tall; breasts began to develop at 14 years, pubic hair at 16:6 years. The standard age for beginning development of the breasts "Brustknospe" is given as 11–13 years (Lipiec, M., 1907).

Table 15 summarizes some findings for these three microsomatics. Lillie S. was the shortest, −25 cm. from control stature at 12 years and −33 cm. at 16 years. Anna W. was 26 cm. short at 12 years and 27 cm. short at 16 years. Marion C. was 20 cm. short at 12 years and 21 cm. short at 16 years. None caught up absolutely in stature in passing from 12 to 16 years and the shortest, Lillie, fell 8 cm. behind during 4 years.

The upper and lower limbs are seen to be extremely short; the most so in the case of Lillie S., where the upper limbs are 14 cm. shorter than control and the lower limbs 20 cm. shorter. In general the deviation from normal decreases with age during the 4-year period. The humerus deviates more than the radius—because it is longer. But though the shank is shorter than the thigh it sometimes deviates the more at 16 years.

In relation to stature at maturity Lillie S.'s

upper limbs are the longest of the three. In relation to sitting height the upper limb is not very aberrant, which indicates slight deviation from the control *ratios*, in contrast to the chondrodystrophics. Likewise the ratios of the upper and lower limbs to each other and of limb segments do not deviate far from standard. Foot length of the microsomatics is, naturally, small and tends, with age, to fall behind the control, perhaps because the foot does not have to carry absolutely as much weight in microsomatics as in control children. However, hand length in relation to stature remains like the control and changes little with age.

Reviewing all the deviations it appears that the three microsomatics differ among themselves and that Lillie S. approaches nearer than the others to the chondrodystrophic type. This finding supports the view of Maas referred to above, that the different categories of dwarfism are not each due to a different distinct cause but that various degrees of activity of different causes are responsible.

B. Acromegaly

This hyperpituitary condition so well described by Cushing (1917: 249–253) was found in one of the boys studied (M. G.), of Italian origin, born in New York City, October 7, 1914. He had suffered an attack of meningoencephalitis. He had an I.Q. of about 50. He was first measured by me in September 1924 when he was 9:11 years old. He was observed and measured annually until 17:6 years of age. In September 1924 he was 152.6 cm. tall (control U. S., 137

TABLE 15

DEVIATION OF MICROSOMATICS FROM CONTROL AT 12 AND 16 YEARS

	Marion C.		Lillie S.		Anna W.	
	12 yrs.	16 yrs.	12 yrs.	16 yrs.	12 yrs.	16 yrs.
Gross upper limb length	−11.3 cm.	−10.1 cm.	−5.0 cm.	−13.7 cm.	−12.8 cm.	−8.8 cm.
Stature	−2.0 cm.	−2.1 cm.	−2.5 cm.	−3.3 cm.	−2.6 cm.	−2.7 cm.
Humerus length	−4.5 cm.	−4.8 cm.	−6.4 cm.	−6.6 cm.	−4.8 cm.	−4.4 cm.
Radius length	−4.2 cm.	−3.6 cm.	−5.1 cm.	−5.8 cm.	−3.5 cm.	−4.1 cm.
Gross lower limb length	−17.1 cm.	−12.4 cm.	−20.8 cm.	−20.2 cm.	−16.3 cm.	−11.8 cm.
Thigh length	−10.0 cm.	−3.6 cm.	−9.6 cm.	−9.2 cm.	−6.2 cm.	−4.8 cm.
Shank length	−7.9 cm.	−6.9 cm.	−8.7 cm.	−10.0 cm.	—	—
$\frac{\text{Gross upper limb length}}{\text{Stature}}$	−28%	+4%	+10%	+25%	+7%	+23%
$\frac{\text{Gross upper limb length}}{\text{Sitting height}}$	−6%	+2.5%	−7.2%	+2.3%	0	+9.0%
$\frac{\text{Upper limb length}}{\text{Lower limb length}}$	+0.7%	−2.0%	+2.9%	+2.3%	+0.8%	−1.2%
$\frac{\text{Humerus}}{\text{Net upper limb length}}$	+0.9%	+1.6%	+1.5%	+0.6%	−0.1%	0
$\frac{\text{Humerus length}}{\text{Femur length}}$	+1.3%	−3.9%	+3.3%	+0.6%	0	−0.3%
$\frac{\text{Radius length}}{\text{Humerus length}}$	−1.5%	−2.6%	−1.4%	−2.5%	−0.6%	0
$\frac{\text{Thigh length}}{\text{Net lower limb length}}$	+2.1%	+3.2%	+0.1%	+1.8%	—	+1.7%
$\frac{\text{Shank length}}{\text{Thigh length}}$	−3.3%	−10.2%	−0.7%	−5.9%	—	−5.4%
Hand length	−3.2 cm.	−3.2 cm.	−2.7 cm.	−3.8 cm.	−4.1 cm.	−3.4 cm.
Hand width	−1.1 cm.	−1.4 cm.	−0.5 cm.	−1.1 cm.	−1.5 cm.	−1.6 cm.
$\frac{\text{Hand length}}{\text{Stature}}$	0	0	+0.5%	+0.7%	+0.2%	+0.4%
Foot length	−2.7 cm.	−2.2 cm.	−4.0 cm.	−5.2 cm.	−4.5 cm.	−5.2 cm.

cm.) and weighed 34.25 kg. (control U. S., 30.3 kg.); in April 1932 he was 186.5 cm. tall (control U. S., 172 cm.) and weighed 68.38 kg. (control U. S., 58.3 kg.). His adolescent spurt of growth occurred during his sixteenth year, when he grew in stature at the rate of 7.9 cm. p. a.

TABLE 16

THE DEPARTURE IN MM. OF VARIOUS DIMENSIONS OF M. G. FROM THE CONTROL U. S. MALE, AT AGES 12 AND 16 YEARS

	12 yrs.	16 yrs.
Gross upper limb length	+9.5 cm.	+6.0 cm.
Net upper limb length	+3.5 cm.	+5.0 cm.
Humerus length	+3.4 cm.	+1.7 cm.
Radius length	+1.4 cm.	+3.2 cm.
Gross lower limb length	+14.0 cm.	+14.0 cm.
Thigh length	+7.1 cm.	+9.9 cm.
Shank length	+6.8 cm.	+6.3 cm.
Hand length	+1.9 cm.	+1.7 cm.
Foot length	—	+4.4 cm.
$\frac{\text{Gross upper limb length}}{\text{Stature}}$	−16%	+1%
$\frac{\text{Gross upper limb length}}{\text{Sitting height}}$	+5%	−8%
$\frac{\text{Gross upper limb length}}{\text{Gross lower limb length}}$	−2.3%	−4.6%
$\frac{\text{Net upper limb length}}{\text{Net lower limb length}}$	−4.0%	−4.6%

From this table it appears that the gross lower limb is more excessively elongated than the gross upper limb and that the foot's length is more aberrant than the hand's. The humerus, at 12 years, is more aberrant than the radius, but by 16 years the humerus seems to have grown more slowly so as to be less extremely elongated than the radius.

In acromegaly just as in chondrodystrophy the lower limb's growth in length is most affected. Also in both abnormal growths the femur is more affected than the radius; but, of course, the two types of abnormal growth are affected in opposite directions. This suggests strongly that the same growth-stimulating organ is involved, in one case working in excess, in the other in deficiency. Also the whole limb and its constituent long bones are affected in proportion to their standard length.

A summary of the dimensions and proportions of the various ages, sexes and endocrine cases is given in table 17. It will be noted that despite the difference in size of grown control boys and girls the proportion of the dimensions is fairly constant, the dimensions of girls being about 93 per cent that of boys. Noteworthy is the great length of the bones in the acromegalic and the shortness of the bones in microsomics, cretins, and Mongoloids and even more in the chondrodystrophics. The shank is especially short in the cretin.

In the comparison of various percentage ratios with the standard ratio the acromegalics are most extreme. In relation to sitting suprasternal, gross lower limb length is 20 per cent greater than control, and gross upper limb length 17 per cent greater. Brachial index is 4 per cent greater and femoro-humeral index 11 per cent less.

In contrast, the chondrodystrophic has a femoro-humeral index 17 per cent greater than control, radio-tibial and intermembral indices 14 per cent greater, but gross lower limb length to sitting suprasternal height is 10 per cent less.

On the other hand the microsomics have indices that do not depart greatly (9 greater to 4 less) from the control; and in the microcephalic also the limb ratios are not far from control.

The Mongoloids have exceptionally short limbs in relation to trunk, and a relatively large radio-tibial index because of a very short shank, though not so short as that of the cretin.

On the basis of these figures alone one might say the ratios of the chondrodystrophics are the most aberrant of all, having an average departure of 10.4 points, and the acromegalics come next with an average departure of 9.3. The microsomics are the least aberrant, having an average departure of only 3.7 points from control.

11. LONG BONE GROWTH, NEGATIVE GROWTH, SELF-SCULPTURING AND RESPONSIVENESS TO STRESSES OF BONE

Growth of bone.—The long bones of the limb illustrate in simplest fashion the linear growth of bones. As is well known the bone is preformed in cartilage. Ossification starts in the diaphysis of the various bones at different ages, but around the seventh week of embryonic life—in the femur and in the humerus at the forty-second day; in the radius and ulna during the seventh embryonic week. Growth of cartilage continues at the epiphyseal ends until about the twentieth year when union of epiphysis with shaft occurs and growth ceases. The precise age of cessation of growth varies. Some children whose growth has been retarded may continue to grow some months longer.

TABLE 17

COMPARISON OF DIMENSIONS AND RATIOS

(Numbers in parenthesis are percentage ratios of the value standing just above to the standard value of the same sex. Owing to the small numbers involved in the endocrine cases that part of the table is only approximately significant.)

	Gross upper limb length, cm.	Gross lower limb length, cm.	Humerus length, cm.	Radius length, cm.	Thigh length, cm.	Shank length, cm.
White U. S. males, birth	20	21	8	7	11	8
White U. S. males, 10 yrs.	58	70	23	19	33	29
(All groups below are taken at age 17 yrs.)						
Control U. S. females	73	85	29	23	42.5	37
Control U. S. males	78	91	31	25	45.5	38
17 yr. f./m. percentage ratio	94	93	94	92	93	93
Cretin, 1 f.	47	56	19	16	28	23
	(64)	(66)	(66)	(70)	(66)	(62)
Chondrodystrophic, 2 f.	51	60	16	14	30	27
	(70)	(71)	(55)	(61)	(71)	(73)
Mongoloid, 50 m.	65	77	26	20	36	33
	(83)	(85)	(84)	(80)	(79)	(87)
Microsomic, 2 f.	59	70	25	21	39	28
	(81)	(82)	(86)	(91)	(92)	(76)
Acromegalic, 1 m.	83	108	35	28	54	46
	(107)	(120)	(113)	(112)	(119)	(121)
Microcephalic, 1 m.	69	84	28	24	41	36
	(89)	(92)	(93)	(96)	(90)	(95)

	Gross upper limb length / Sitting suprasternal ht. ×100	Intermembral Index	Femoro-Humeral Index	Brachial Index	Radio-Tibiale Index	Gross lower limb length / Sitting suprasternal ht. ×100	Shank length / Thigh length ×100
White U. S. males, birth	100	99	78	90	85	102	84
White U. S. males, 10 yrs.	132	64	72	80	65	156	89
(All groups below are taken at age 17 yrs.)							
Control U. S. females	120	82.5	69.8	80	61.5	147	88
Control U. S. males	130	81.5	69.5	80	65.5	165	87
17 yr. f./m. percentage ratio	92	101	101	100	94	89	101
Cretin, 1 f.	113	84	70	82	68	135	82
	(94)	(102)	(100)	(102)	(111)	(92)	(93)
Chondrodystrophic, 2 f.	123	94	82	75	70	132	80
	(103)	(114)	(117)	(94)	(114)	(90)	(91)
Mongoloid, 50 m.	122	82	68	76	68	149	89
	(94)	(101)	(98)	(95)	(104)	(90)	(103)
Microsomic, 2 f.	130	83	69	77	65	153	87
	(109)	(101)	(99)	(96)	(106)	(104)	(99)
Acromegalic, 1 m.	152	77	62	83	62	198	84
	(117)	(95)	(89)	(104)	(95)	(120)	(97)
Microcephalic, 1 m.	141	82	70	83	67	175	89
	(108)	(101)	(100)	(104)	(102)	(106)	(102)

Growth in length of the long bones is exclusively at the ends. This has been determined by the drilling of holes, inserting of pins or shot, in young mammals, at measured distances apart and then remeasuring the distances after the lapse of months or years (Hales, 1727; Duhamel, 1743). Such observations and experiments have been repeated down to the latest studies of J. D. and M. E. Bisgard (1935).

Growth of the long bones is, then, at the two ends and is accompanied by the multiplication of the columns of cartilage cells and their activity in producing the collagen and water of the formed substance.

While these secretions are proceeding constantly they receive a special acceleration by the presence of hormones. This has been beautifully shown by the studies of Freud, Levie, and Kroon (1939). Working with rats they found that after hypophysectomy growth in length (of

vertebræ of tail) ceased. Specifically there was no longer a direct growth of cartilage where cells were distributed irregularly. There was an increase in amount of matrix and eventual closure of the growth zone of the epiphyseal region by a bony plate. Growth was restored by the intraperitoneal administration of hypophyseal hormone. The premature closure of the epiphysis (which occurs after hypophysectomy) can be hindered by the growth hormone. This "growth hormone" reacts upon the proliferating cartilage; it is more properly called the chondrotrophic hormone.

Shortening of long bone at puberty.—The curves of length of individual long bones tend to become asymptotic as maturity is reached. In some girls, however, there is clearly a reduction in length at puberty. Examples are seen for the humerus in figure 10, curve 7; figure 11, curves 1, 3, and 7 and in figure 12, curves 1, 2, and 6; and for the femur. This shortening is not found in the curves of the radius (fig. 15), nor in the tibia, so far as analyzed. It is shown in figure 18, curve no. 16 (gross lower limb). It is apparently shown in some *mass* curves, such as figures 1, 4, 8, 16, 19. It is found once in twin boys (femur); but these boys are strikingly effeminate in behavior.

This shortening seems to be a biological phenomenon, and not at all impossible in living, constantly changing, developing bone.

There is some good evidence that the long bones constitute a reservoir of calcium salts that may be drawn upon recurrently. Thus as Riddle and Reinhart (1926) have shown the serum calcium of the female pigeon is regularly and normally increased by more than 100 per cent whenever her ovary functions. In mammals it has been shown by Talbot (1939) that when estradiol is injected into immature rats the skeletal age of the females (but not of the males) was advanced. Thus calcium seems to be withdrawn from the bones under some circumstances and added to the bone (with possible change in form) under other circumstances. Possibly the epiphyseal union which seems to be accelerated, in dogs, by estrogen (Tausk and de Fremery, 1935) may result in a reduction of the distance from one end of the long bone to the other. This field is insufficiently worked, but enough is known to leave it not unreasonable that long bones may shorten a few millimeters in connection with the fundamental changes, especially in the female, at puberty.

Self-sculpturing of bone.—As the bone as a whole elongates, the region of the broad head is left behind and becomes incorporated in the slender column while new material is formed at the broad advancing head. This process makes necessary the absorption of the margins of the part of the head that has been left behind and the reconstruction of this part of the bone to form part of the column (pl. 2, *C–I*). There is thus a continuous process of adding new bone on the broad surface of the ends and later resorbing and remolding the head of the bone as it is transformed into the column. The ends of the finger muscles which are attached to the periosteum, or membrane covering the head, apparently move constantly to new positions as new material is formed at the ends of the bone.

Since preparing plate 2 I find that Brash (1934) published a figure giving the superposition of three radiograms of the end of the femur of the same child during three successive years. This shows plainly the phenomenon of self-sculpturing that I have referred to. However, I have the advantage of having been provided, by the courtesy of Letchworth Village, with a series of four to six radiograms of the bones of the hands taken at their birth-month for each of fifty boys and fifty girls. From a study of these successive radiograms the phenomenon to which Brash called attention proves to be practically a universal one. The only exceptions have been a few cases of exceedingly thick metacarpal bones in which the comparatively narrow head seemed to be absorbed without reduction in the widening column as the column elongated.

The details of the form of the bones are under the control of hereditary factors. Thus plate 2, *A* shows, superimposed, the outlines of the hand bones of the G. twins taken at the same time. The details of form are often reproduced with great precision so that the outlines would coincide if superimposed. The resemblance is the more striking if a comparison of the same bone at the same age is made on a number of different individuals; as in plate 2, *C–I*. Note the slenderness of *E* and the grossness of the *F*, as compared with *H* and *I*, corresponding bones of the twins. When we recall that the bones are being constantly reformed at the ends as the ends become transformed into the shaft the eventual similarity of the sculpturing in twins, despite the great variety of possibilities, offers food for thought.

These various observations indicate not only

the great plasticity of bone but the presence of form-determining factors in the bone which mold it to an adaptive form. Each bone, though all are provided with the same hormones, grows in its specific way to produce its particular form.

Responsiveness of bone.—In some contrast to the self-sculpturing of bone chiefly due to internal form-determining factors (but certainly modifiable by the external factor of use) is the determination of the internal structure of bone. It has been known for seventy years that the upper end of the thigh bone shows, on a longitudinal median section, a system of trabeculae such as are fit, with least weight of materials, to support superimposed pressures. Even the interior of the heel bone shows numerous bony rods that resist the blows and pulls upon it. In the absence of such molar agents, as in the case of cripples who have never walked, the general form is modified and the internal structure shows, in the roentgenogram, no such dense rods and plates as in the normal child. In the case of a child who walked till 10 years of age and has been bed-ridden since, the form of the bone is nearly typical but the inner rods are much reduced in number and density (Davenport, 1941). The bone shows itself highly responsive to use and reacts to the agents that impinge upon it.

12. MAN'S PLACE AMONG THE MAMMALS

Man's place among the mammals is determined first by his origin and secondly by his present functional position. By a study of the changes in the bones of the limb from intrauterine to adult stages we may trace, in outline, his later phylogenetic history.

Before birth the percentage ratio of upper limb length to sitting suprasternal height increases from 80 (at 10 weeks) to 110 at 20 weeks and then decreases to 95 at 6 months, p. p. The anthropoid apes indeed show a slight reduction of this ratio in infancy of 1 to 5 points (Schultz, 1926*a*: 488). The high point reached *in utero* suggests a development along anthropoid lines, followed by a marked reduction of the ratio (rapid growth of sitting suprasternal height) in the infant from which partial recovery is later made, though the ratio in adult man is still 25 points or more less than in the adult gibbon.

Very different is the age change of percentage ratio of lower limb length to sitting suprasternal height. There is an infantile depression but this is rapidly followed by a great increase in the ratio of 64 points, unlike anything found in other Primates. The great addition in the relative growth of the human lower limb after a low ratio at birth follows again the phylogenetic path.

Again, the forearm/"arm" percentage ratio which is about 73 at 6 weeks rises to 90 at birth as does that of Hylobates (Schultz, 1926*a*: 490). But while the ratio continues to increase in *Hylobates* to 114 in the adult it decreases in *Homo Europaeus* to 80 or less.

The shank/thigh length percentage ratio begins very low in *Homo;* even below that in *Hylobates*. In both species the adult percentage ratio is about 88 (fig. 48).

In respect to the intermembral index (upper limb length/lower limb length percentage ratio), I may quote from my 1935 paper.

> The intermembral index of the anthropoid apes is a very large one. In the case of the adult gibbon it is in the neighborhood of *135* and even the young have a large index. Indeed, apparently in most Primates the arm develops at first faster than the leg, but before the middle of gestation the arm begins to lag behind the leg. At any rate, the human fetal proportions are like those of the young gibbon. The infant gorilla has an intermembral index of about 125 and this decreases in the adult to about 117. The arm of the infant, therefore, begins to grow like the arm of other Primates: long at first and then relatively shorter; but whereas in most other Primates the arm remains longer than the leg, in the child the leg early begins to grow so much more rapidly that at birth the arm is already shorter than the leg.
>
> To the comparative mammalogist the high intermembral index of the human fetus indicates a derivation of *Homo* from types which had strikingly long arms. The human child, in relative length of arm and leg, follows first the ordinary anthropoid path but goes beyond the other anthropoids in relative arm reduction. In view of the fact that *Tarsius* and *Lemur* have a relatively low intermembral index the high index in the human child would suggest immediate origin from brachiating rather than quadrupedal ancestors.

The percentage ratio of the length of the "arm" to the net upper limb length is small in the gibbon (47), as compared with man (56). But in gorilla and chimpanzee the ratio is much like that of man. The differentiation curve for man shows at birth a low percentage ratio—53. Apparently the gibbon retains that low ratio and gains a lower one while in the other anthro-

TABLE 18

MEAN LENGTHS OF FEMUR AND HUMERUS AND PERCENTAGE RATIOS BETWEEN THEM AND OTHER LONG BONES FOR 4 SPECIES OF ANTHROPOID APES: *Hylobates hoolock* (GIBBON), *Pongo abilu* (ORANG-UTAN), *Pan pygmaeus* (CHIMPANZEE), AND *Gorilla gorilla berinzei* (GORILLA). LENGTHS AS FOLLOWS: A, UPPER LIMB; F, FEMUR; H, HUMERUS; L, LOWER LIMB; N, NET; R, RADIUS; T, TIBIA; f, FREQUENCY.

	Gibbon		Orang-utan			Chimpanzee			Gorilla		
	smallest	largest	sm.	med.	larg.	sm.	med.	larg.	sm.	med.	larg.
f	(10)	(3)	(9)	(27)	(8)	(10)	(13)	(7)	(3)	(10)	(8)
F	188	212	110	238	294	190	283	304	170	335	375
$\frac{H}{NA}$	48	48	51	50.5	50	52	52	53	54	55	56
$\frac{R}{H}$	109	107	96	99	102	92	94	88	86	83	80
H	211	256	139	313	370	210	285	323	195	381	445
$\frac{F}{NL}$	54	54	54	54	54	54	54	55	54.5	55	55
$\frac{NA}{NL}$	130	133	133	139.5	136	110	105	109	117	115	117
$\frac{H}{F}$	116	116	127	132	126	107	101	106	115	114	119
$\frac{R}{T}$	146	152	140	149	148	115	110	106	119	116	114
$\frac{T}{F}$	86	86	88	88	87	83	86	84	83	81	83

poids the ratio increases. It seems probable that the low ratio of the gibbon is a special adaptation to its extraordinary brachiating habits and that *Homo's* ancestors did not pass through that extreme phase.

On the whole the changes of proportions in the limbs during development seem to indicate man's origin from ancestors with relatively longer upper limb, shorter lower limb and longer forearm.

The functional place of man among the mammals has been referred to on several of the preceding pages: his low net intermembral index (66) places him between the hare (71) and the jumping mouse (55) as a jumping animal. On account of his low humerus/femur percentage ratio of 69 he is placed between the running and jumping animals and far removed from the brachiating apes. In relation of radius to humerus length, which in average adult European males has a percentage ratio of 80, man is close to the gorilla and to some burrowing mammals and swimmers. In his low index (66) of the radius/tibia length man is again allied to the swimmers and jumpers. The mean crural index (shank length/thigh length) in adult Europeans is 84. This places man not far from the bear (85) and agouti (81) and indeed in the class of runners. Summing up these conclusions as to man's functional place among mammals it appears that he has passed from the group of brachiating animals and come to join the groups of runners, jumpers, swimmers, and even burrowers. He has evolved to a point where he is better fitted for land and water than for trees.

This may be a suitable place to present a general table of the relation in length of the long bones of the limbs in the four genera of anthropoid apes (table 18). The data given in table 18 are based on dry bones from the American Museum of Natural History and the United States National Museum, measured by myself.

It will be observed that the individuals are of different sizes. The arrangement in the table is made on the basis of length of femur. Individuals with about the same length of femur are grouped and averages are given. Of course the larger specimens are not necessarily older (as is true also in *Homo*) but the very smallest were obviously immature on account of ununited epiphyses.

Referring to the radius/humerus length ratio, this is approximately constant in the chimpanzee, decreases with age in the gorilla, in-

creases with age in the orang and is large in the gibbon. This suggests that the ancestral, pre-anthropoid, form had a youthful index of not far from 90; that the chimpanzee has retained this primitive condition; that in the gorilla the ratio tends to diminish as it becomes adult as occurs in *Homo;* that in the orang and still more markedly in the gibbon the brachial index becomes adapted during postnatal development to a condition of long radius.

13. DIMENSIONS AND PROPORTIONS OF LIMBS IN RELATION TO FUNCTION

In view of the great variation in the length of the limbs and in the proportional length of the segments of the limbs, the query must often arise if there is any relation between success in their use in various athletic operations and these physical dimensions and relations.

The most thoroughgoing study of this subject has been made by Bach (1926, 1927, 1929, 1930). He finds that the "Mehrkämpfer" (i. e. runners, jumpers, walkers, etc.) are characterized by longer limbs than the average; gymnasts by long trunks; wrestlers by breadth of trunk. He finds (1930: 96) that the high jump increases with length of lower limb. Beall (1939: 46, 48, 49) finds that, in college girls, success in basket ball is associated with longer upper limb, larger foot area, and broader shoulders; in swimming with greater weight and breadth; in tennis playing with greater stature and lower limb length; in dancing with short slender form and limbs. "*Relative* dimensions are of no statistical significance in comparing successful with unsuccessful groups in each activity nor in comparing one activity with another." That is, long lower limbs will win the running race whether the rest of the body is long or short. To these studies I have been able to make no contribution.

IX. SUMMARY OF CONCLUSIONS

The upper limb from 3 to 12 years increases about 2.5 cm. p. a. in length with a spurt at adolescence. It is shorter in girls than boys, very short in Mongoloid dwarfs, long in Negroes.

The speed of growth of the length of the "arm" (humerus) decreases from midgestation. The growth curves are often wavy, periods of rapid growth alternating with periods of slow growth. The length at puberty may even become shorter in girls but rarely in boys. In ateliotics growth in length of "arm" may continue beyond the age of cessation in standard-sized children.

The rate of growth of the radius, which was almost as fast as that of the humerus before birth, slows down more than the humerus after birth so that in the adult the length of the radius is only three-fourths that of the humerus.

The length of the lower limb is measured by the method of Jazuta.[8] In intra-uterine life growing geometrically, the limb after birth slows down, with accelerations at 9 and 14 years. The lower limbs of girls are longer than those of boys until the boys' adolescent spurt. Races of mankind differ greatly in lower limb length, Negroes having the longest. The acromegalics have long limbs; Mongoloid dwarfs short limbs; monozygotic twins have lower limbs of similar length. The length of the femur increases prenatally, on the average half a millimeter per day.

The shank at birth is about 16 per cent shorter than the femur, and through slow but proportionate growth retains that ratio to adulthood. It is longer in girls than boys till the adolescent spurt of boys. The tibia is less shortened on standing than the femur. In Negroes the excessive length of shank is more marked than that of the thigh.

The upper limb grows much faster than stature before birth, slightly faster after infancy. But between 6 and 8 months its relative growth diminishes because of the rapid growth of the lower limbs and stature. Growth of upper limb in relation to sitting height is much more rapid than in relation to stature, because of the elimination of the length of the rapidly growing lower limbs from the denominator of the ratio. This ratio is very alike in the two sexes until the onset of adolescence, when the curve for girls falls below that for boys. In males the differentiation curve never slopes downward with age as it does in the chimpanzee. In relation to sitting suprasternal height the upper limb grows rapidly but does not always follow the "law of relative growth."

The intermembral index decreases with age after 2 years; the curve is irregular before that. Girls have a smaller index than boys until puberty. Mongoloid dwarfs and an acromegalic also have a small index but Negroes and chondrodystrophics have a high one.

[8] Different methods of measuring the lower limb are described by Gray, 1941: 375 (see also page 386 of this text).

The net intermembral index decreases with age after 1½ years p. p.; it is smaller for girls than boys, large in idiots, Negroes and chondrodystrophics, and small in Mongoloid dwarfs. In boys the index reaches a minimum just before the adolescent spurt.

The mass curve of the percentage ratio of "arm" length to gross upper limb length is constant after 6 years; but individuals show waves of change in the ratio, and these are still more striking when the relation of growth of "arm" length to that of net arm length is plotted. Similarly the humerus and stature as a whole grow at very different rates at different times.

The femoro-humeral index shows a slow reduction after 3 years p. p. The great development in length of the femur places *Homo* between the running and the jumping animals, far removed from the brachiating apes.

The brachial index shows the wavy type of curve, i. e., rhythmic growth. His index of 80 places man among the swimmers and ground mammals.

His radio-tibial index places man far below such brachiating species as gorilla and chimpanzee; and nearer to the swimmers and burrowers.

The percentage ratio of length of the lower limb to sitting suprasternal height shows an increase from 100 at birth to 165 at maturity. At maturity the percentage ratio ranges from 133 to 200; being highest in acromegalics. In individual curves of change with age a maximum is achieved in boys at the time of the adolescent spurt.

The curve of mean thigh length in relation to net lower limb length though fairly smooth is made up of individual curves of the wavy type, apparently owing to growth rhythms. Though the individual curves are complex they run closely parallel in monozygotic twins. Shank length in relation to lower limb length shows similar phenomena. The individual curves of crural index change with age and show the same strong waves.

The ratio of length of the shank to stature increases with age. The ratio is greater in girls than in boys to adolescence.

During development, span increases faster than stature, reaching, on the average, 100 per cent of stature at about 8 years and 102.5 per cent at 16 years. The ratio is usually smaller in girls than in boys; and much greater in Negroes than in Whites. A minimum is frequently passed through by the curves of boys at 12 or 13 years of age.

Arm girth is larger in girls than boys and shows great individual variation. The minimum girth of forearm changes very little during childhood (3 to 8 years).

Thigh girth varies with nutrition, general health, exercise and especially with rate of linear growth of the femur. The same holds true, but less strikingly, for calf girth. Ankle girth, on the other hand, is less modified by use.

The girth of the "arm" is likewise inversely related to speed of growth of the humerus. The faster a long bone grows the more the girth of the limb is reduced. As the limb is more used with advancing age the ratio of minimum to maximum girth decreases rapidly.

The hand grows longitudinally and laterally in smoother fashion than the long bones of the limbs. The hand index is large in both human and chimpanzee fetuses but the chimpanzee acquires a much slenderer hand than *Homo*. The Negro has a long, slender hand; the Mongoloid dwarf a stubby one. Hand area is smaller in girls than in boys, except temporarily at the time of female adolescence. While hand area differs widely in individuals, in twins the area is nearly the same. Hand length decreases relative to limb length during development. In dwarfs the growth of the bones of the hand seems to be less affected than that of the longer bones of the limbs. Hand length in relation to forearm length and to stature undergoes great fluctuations in childhood.

The foot grows rapidly in length before birth and very slowly between 4 and 8 years. The foot index is highest at birth as in the chimpanzee, but does not as in that ape decrease with age. Thus early recapitulation is suggested. The height from the floor of the malleolus (sphyrion) and that of the naviculare continues to increase to maturity, reaching 4 and 2 cm. respectively. The arch index tends to increase with age, but in a few cases the age curve slopes downward at the later years showing "fallen arches."

The interval between the radial stylion, with arms hanging, and the tibiale varies at 16 years, from 26 to 38 cm. It tends to increase till adolescence and may even decrease after puberty. The curve of change of this interval in terms of stature is concave upward with a minimum at 7 years.

The gluteal masses (buttocks) which hold the body erect tend to increase rapidly in depth until

8 months p. p. The maximum index is reached, in boys, at about 11 years. In relation to stature the gluteal index diminishes till about 9 years, and then, in boys, the index (depth/height) increases for some years.

During postnatal development of the average child the scapulae increase in length *pari passu* with increase in size and activity of the upper limb. Also the angle that the scapulae make with the frontal plane tends to decrease from the quadrupedal position to the mature position, approaching parallelism with that plane. This is again apparently an example of recapitulation.

From a comparison of differences between successive measurements of the same dimension and of purely chance errors a guide is obtained as to how to plot the curve so as to approach most closely to the true value.

The importance of studying separately growth and differentiation as factors in development is emphasized and a method of expressing growth in Ångstrom units per second illustrated.

A real rhythm of growth of the long bones, first announced by Godin, is found everywhere—especially in following the growth of individuals. All bone growth is probably rhythmic or at least varies in rate from time to time. Such variability in growth rate of a given tissue is probably a wide-spread phenomena.

Attention is called to the obvious but sometimes overlooked fact that the dimensions and proportions of a child are a function of gravity.

Using the average boy as control, girls have shorter upper and lower limbs and especially longer tibiae; they have longer lower limbs in relation to stature; the thigh forms a smaller fraction of the limb and the tibia conversely a larger one. The hand index is smaller.

The control boys rank highest in limb length, are low in intermembral index, and have the greater relative upper limb index and upper to lower limb length ratio until surpassed by the idiots after adolescence. They have the "arm" forming a greater portion of the upper limb than in other sexes, have a low crural index, high relative span and slender hand.

In relation to Whites the Negroes have, first, long upper limbs and, secondly, long lower limbs. The distal segments are especially long, as in the gibbon.

The limbs grow in length in babies in diverse fashion, and not in accordance with a universal "law of growth." The growth of each part is influenced by that of neighboring parts. Also in youth growth is proceeding in different parts at different rates and hence producing adult organs of different size and shape.

In fraternities the growth and differentiation curves sometimes differ markedly; but even more striking is their frequent parallelism and the evidence of segregation of such curves inside of one fraternity, giving evidence of genetic control. The evidence of such control is most striking in the case of monozygotic twins. The degree of agreement in the curves of twins is, however, different in the different dimensions and ratios. Thigh and shank length and hand length especially show deviations from parallelism. Of the ratios, those involving gross upper limb length often show non-parallelism.

Among dwarfs the cretins are characterized especially by short lower limbs and by small thigh girth. The length of the thigh seems to be less affected by thyroid therapy than the length of the "arm." In the chondrodystrophic man the lower limbs are absolutely more shortened than the upper; the net upper limb length/sitting height percentage ratio is 33 below controls while the hand/forearm length percentage ratio is 27 above controls. In Mongoloids the thigh is slightly more shortened than the humerus, and the shank than the radius. The reduction in length of the long bones is less than in the chondrodystrophic dwarf.

The various microsomatics studied depart from control in different fashion, supporting the view that the different categories of dwarfism are not due each to a different distinct cause but that various degrees of activity of different causes are responsible.

In acromegaly the growth in length of the lower limb is most affected, and the femur more than the radius. This suggests that the same growth-stimulating organ is involved in chondrodystrophy and acromegaly; in one case working in deficiency, in the other in excess.

The growing long bones are exceedingly responsive to mechanical stimuli both internal and external and both in general form and in structure. The near identity of the form of the hand bones in monozygotic twins measures the control of hereditary factors. The long bones may shorten slightly in girls at puberty.

Functionally, man has passed from the groups of brachiating animals and come to join the groups of runners, jumpers, swimmers, and even burrowers. He has evolved to a point where he is better fitted for land and water than for the

trees. In his ontogeny he shows, at every step, evidence of the path of his evolution.

X. BIBLIOGRAPHY

AKIBA, T. 1924. *Folia Anatom. Japonica* **2**: 189–219.

APPLETON, V. B. 1928. *Amer. Jour. Phys. Anthrop.* **11**: 473–500.

BACH, F. 1926. *Zeit. mensch. Vererb. KonstLehre.* **12**: 469–523.

—— 1927. *Ibid.* **13**: 219–240.

—— 1929. *Deutsche Turn-Zeitung*, Nos. 46–48.

—— 1930. Leitfaden zu anthropometrischen Sporttypenuntersuchungen und deren statistischer Auswertung. *Ärtzliche Rundschau.*

BARDEEN, C. R. 1905. *Amer. Jour. Anat.* **4**: 265–302.

BAYER, L. M., AND H. GRAY. 1933. *Amer. Jour. Phys. Anthrop.* **17**: 379–415.

BEALL, E. 1939. The relation of various anthropometric measurements of selected college women to success in certain physical activities. Teachers College, Columbia Univ. *Contrib. to Education*, No. 774.

BISGARD, J. D., AND M. E. BISGARD. 1935. *Arch. Surg.* **31**: 568–578.

BRASH, J. C. 1934. *Edinburgh Med. Jour.* **41**: 305–319; 363–387.

BREZINA, E., AND V. LEBZELTER. 1925. *Zeit. mensch. Vererb. KonstLehre* **10**: 381–389.

BROCA, P. 1862. *Bull. Anthrop. Soc. Paris.* **3**: 162–172.

CAMPBELL, T. D., AND C. J. HACKETT. 1927. *Trans. Roy. Soc. So. Australia* **51**: 65–75.

CIPRIANI, L. 1932. *Archivio Antrop. Etnol.* **60–61**: 136–286.

COLLINS, H. B., JR. 1928. *Amer. Jour. Phys. Anthrop.* **11**: 353–355.

CORREIA, A. C. G. D.S. 1928. Les lusos-descendants de l'Inde portugaise. Nova Goa.

CUSHING, H. 1912. The pituitary body and its disorders. Lippincott.

CZEKANOWSKI, J. 1910. *KorrBl. deuts. Ges. Anthrop. Ethnol. Urgeschichte* **41**: 101–109.

DAFFNER, F. 1902. Das Wachstum des Menschen. Leipzig, Engelmann.

DAVENPORT, C. B. 1926. *Amer. Jour. Phys. Anthrop.* **9** (2): 217–220.

—— 1927. Guide to physical anthropometry and anthroposcopy. Eugenics Assn., Cold Spring Harbor, N. Y.

—— 1932. *Amer. Jour. Phys. Anthrop.* **17**: 167–211.

—— 1933. *Ibid.*: 333–353.

—— 1934. *Proc. Nat. Acad. Sci.* **20**: 359–364.

—— 1935. *Jour. Mammalogy* **16**: 291–296.

—— 1937. *Growth* **1**: 279–283.

—— 1938. *Carnegie Inst. Washington Publ.* No. 496: 275.

—— 1939. *Proc. Amer. Philos. Soc.* **80**: 175–352.

—— 1940. *Ibid.* **83**: 1–215.

—— 1941. *Ibid.* **84**: 65–70.

—— AND A. G. LOVE. 1921. Army Anthropology. Med. Dept. U. S. Army in World War, **15**. Govt. Printing Office.

—— AND GRACE ALLEN. 1925. *Jour. Psycho-asthenics* **29**: 266–286.

—— AND W. W. SWINGLE. 1927. *Jour. Exp. Zool.* **48**: 395.

—— AND M. STEGGERDA. 1929. Race crossing in Jamaica. *Carnegie Inst. Washington Publ.* No. 395.

—— M. STEGGERDA AND W. DRAGER. 1934. *Proc. Amer. Acad. Arts & Sci.* **69**: 267–268.

—— O. J. RENFROE AND W. D. HALLOCK. 1939. *Child Development* **10**: 181–202.

DUHAMEL, H. L. 1743. *Mém. Acad. roy. sci., Paris.*

DUNN, L. C. 1927. *Arch. EntwMech. Org.* **110**: 341–365.

EICKSTEDT, E. V. 1926. *Mitt. Anthrop. Ges. Wien* **56**: 173–183.

FICK, R. 1926. *SitzBer. Preuss. Akad. Wiss. Phys.-math. Kl.* **30**: 417–451.

FLOWER, W. H. 1880. *Jour. Anthrop. Inst.* **9**: 108–135.

FRANCIS, C. C. 1939. *Amer. Jour. Phys. Anthrop.* **25**: 323–331.

FRANZEN, R. 1929. Physical measures of growth and nutrition. *School Health Research Monograph* **2**.

FREUD, J., L. H. LEVIE AND D. B. KROON. 1939. *Jour. Endocrin.* **1**: 56–64.

FRIEDENTHAL, H. 1914. Allgemeine und specielle Physiologie des Menschenwachstums. Berlin, Springer.

GESELL, A., AND H. THOMPSON. 1938. The psychology of early growth. Macmillan.

GODIN, P. 1903. Recherches anthropométriques sur la croissance des diverses parties du corps. Paris, Maloine.

—— 1935. Ditto. 2me édition, completée par les resultats nouveaux concernant l'influence de la puberté sur les accroissements. Paris.

GRAY, H. 1941. *Growth* **5**: 373.

—— AND J. G. AYRES. 1931. Growth in private school children. Univ. of Chicago Press.

GRÜTZNER, G. 1928. *Arch. Klaus-Stift. VererbsForsch.* **3**: 95–218.

GUDERNATSCH, F. 1934. *Cold Spring Harbor Symposia on Quant. Biol.* **2**: 94–106.

GÜNTHER, H. 1938. *Zeit. mensch. Vererb. KonstLehre* **22**: 342–355.

HALES., S. 1727. Vegetable staticks: or, an account of some statistical experiments on the sap in vegetables. London, Innys.

HANNESSON, G. 1925. Körpermasze und Körperproportionen der Isländer. Reykjavik, Iceland.

HARRIS, H. A. 1931. *Amer. Jour. Med. Sci.* **1811**: 1–26.

HENCKEL, K. C. 1929. *Gegenbaurs morph. Jahrb.* **61**: 43–49.

HOESCH-ERNST, L. 1906. Das Schulkind in seine körperlichen und geistigen Entwicklung. Leipzig, Nemmich.

HRDLIČKA, A. 1925. The old Americans. Williams & Wilkins.

—— 1928. *Amer. Jour. Phys. Anthrop.* **12**: 15–33.

JAZUTA, K. Z. 1925. *Arch. Anthrop.* n.f., **20**: 154–156.

JERUMS, N., AND T. M. VITOLS. 1928. *Acta Univ. Latviensis* **18**: 1–99.

KAJAVA, V. 1927. *Ann. Acad. Scient. Fennicae* **27**.

KATO, V. 1938. *Rassenbiol. Untersuch.* **5**: 133–194.

KOGANEI, Y. 1894. *Mitt. med. Fak. Tokio* **2**: 251–404.

KUGLER, E. 1932. *Arch. Klaus-Stift. VererbsForsch.* **6**: 429–577.

LANDAUER, W. 1927. *Arch. EntwMech. Org.* **110**: 195–278.

LIPIEC, M. 1907. *KorrBl. deuts. Ges. Anthrop.* **38**: 175–180.

LUNDBORG, H., AND F. J. LINDERS. 1926. Racial characters of the Swedish nation. Uppsala.

MAAS, O. 1920. *Zeit. ges. Neurol. Psych.* **57**: 196–231.

—— 1923. *Ibid.* **86**: 26–85.

—— 1925. *Ibid.* **95**: 784–808.
MARSELIS, J. G., AND K. TEXLER. 1925. *Zeit. mensch. Vererb. KonstLehre* **11**: 576–638.
MARTIN, RUDOLF. 1928. Lehrbuch der anthropologie. Jena, Fischer.
MIDLO, C. 1934. *Amer. Jour. Phys. Anthrop.* **19**: 337–389.
MINOT, C. S. 1913. Modern problems in biology. Blakiston.
MITCHELL, A. 1876. *Jour. Mental Sci.* **22**: 174.
MOLLISON, T. 1910. *Gegenbauers morph. Jahrb.* **42**: 79–304.
MORTON, D. J. 1922. *Amer. Jour. Phys. Anthrop.* **5**: 305–336.
—— 1924. *Ibid.* **7**: 1–52.
—— 1924. *Jour. Bone and Joint Surgery* **6**: 56–90.
NIGGLI-HÜRLIMANN, B. 1930. *Arch. Klaus-Stift. Vererbs-Forsch.* **5**: 1–215.
OGAWARA, S. 1933. *Zeit. Mensch. Vererb. KonstLehre* **17**: 90–154.
PAPILLAULT, G. 1902. *Bull. Soc. Anthrop. Paris* (5) **3**: 393–526.
PONIATOWSKI, S. 1911. *Arch. Anthrop.* n.f. **11**: 249–279.
RHEIL, A. 1926. *Zeit. Morph. Anthrop.* **26**: 333–358.
RIDDLE, O. 1927. *Proc. Amer. Philos. Soc.* **66**: 497–509.
—— AND W. H. REINHART. 1926. *Amer. Jour. Physiol.* **76**: 660–676.
RODENWALT, E. 1927. Die Mestizen auf Kisar, 2 v. Nederl. v. d. Dienst d. Volkgesondheit f. Nederlandisch-Indie.
ROSE, W. C. 1937. *Science* **86**: 298.
ROTH-LUTRA, K. H. 1930. *Zeit. mensch. Vererb. KonstLehre* **15**: 518–543.
ROTT, A. 1926. *Anthrop. Anzeiger* **3**: 39–45.
SALLER, K. 1930. *Deutsche Rassenkunde* **2**.
—— 1930. *Ibid.* **4**.
SAMESIMA, M. 1938. *Rassenbiol. Untersuch.* **5**: 237–297.
SCAMMON, R. E., AND L. A. CALKINS. 1929. The development and growth of the external dimensions of the human body in the fetal period. Univ. of Minn. Press.
SCHENK, P. 1925. *SitzBer. Ges. Naturw.* No. 2: 7–38.
SCHLAGINHAUFEN, O. 1929. *Arch. Klaus-Stift. Vererbs-Forsch.* **4**: 219–287.
—— 1932–33. *Bull. Schweiz. Ges. Anthrop. Ethnol.*, 29–59.
SCHMALHAUSEN, I. 1925. *Arch. EntwMech. Org.* **105**: 711–717.
SCHREIBER, O. 1928. *Zeit. mensch. Vererb. KonstLehre* **13**: 675–690.
SCHREINER, A. 1932. *Skrifter Norske Videnskaps Akad. Math. Natur. Kl.*, No. 1.
SCHULTZ, A. H. 1925. *Jour. Washington Acad. Sci.* **15**: 247–263.
—— 1926. *Amer. Jour. Phys. Anthrop.* **9**: 65–80.
—— 1926a. *Quart. Rev. Biol.* **1**: 465–521.
—— 1930. *Human Biology* **11**: 303–348.
—— 1931. *Sci. Mon.* **33**: 385–412.
—— 1933. *Amer. Jour. Phys. Anthrop.* **18**: 61–77.
—— 1937. *Human Biology* **9**: 281–328.
—— 1940. *Contrib. to Embryol. Carnegie Inst. Washington*, No. 170.
SCHWARTZ, F. 1910. *Arch. Anthrop.*, n. f. 10.
SHIROKOGOROFF, S. M. 1925. Process of physical growth among the Chinese. Vol. 1. Chinese of Chekiang and Kiangsu. Shanghai, Commercial Press.
—— 1925a. Anthropology of Eastern China and Kwangtung Province. Roy. Asiatic Soc. North China Branch, extra vol. 4. Shanghai, Commercial Press.
SHUTTLEWORTH, G. E. 1909. *Brit. Med. Jour.* **2**: 661–665.
SMITH, T. 1896. *Ibid.*: 616.
STEGGERDA, M., J. CRANE, AND M. D. STEELE. 1929. *Amer. Jour. Phys. Anthrop.* **13**: 189–255.
STRATZ, C. H. 1922. Der Körper des Kindes und seine Pflege für Eltern, Erzieher, Ärzte und Künstler. Stuttgart, Enke.
STUART, H. C., P. HILL AND C. SHAW. 1940. *Monographs Soc. Res. Child Dev.* **5**.
SUTHERLAND, G. A. 1899. *Practitioner* **63** (n.s. v. 10): 632–642.
TALBOT, N. B. 1939. *Endocrinology* **25**: 325–327.
TAUSK, M., AND P. DE FREMERY. 1935. *Acta Brev. Neerl. Physiol.* **5**: 19–22.
THOMAS., A. W. 1934. Colloid chemistry. McGraw-Hill.
TODD, T. W., AND A. LINDALA. 1929. *Amer. Jour. Phys. Anthrop.* **12**: 35–119.
TREDGOLD, A. F. 1914. Mental deficiency. London, Bailliere.
VALLOIS, H. V. 1932. *Compte rendu Soc. Biol., Paris* **110**: 915–917.
VIERORDT, K. H. 1906. Anatomische, physiologische und physikalische Daten und Tabellen zum Gebrauche für Mediziner. Jena, Fischer. (3rd ed.)
VIRCHOW, R. 1879. *Zeit. Ethnol.*: 383–395.
WAGENSEIL, F. 1925. *Zeit. Morph. Anthrop.* **23**: 33–150.
WECHSLER, W. 1939. *Arch. Klaus-Stift. Vererbs Forsch.* **14**: 199–359.
WEINBACH, A. P. 1938. *Growth* **2**: 303–317.
WEISSENBERG, S. 1911. Das Wachstum des Menschen nach Alter, Geschlecht und Rasse. Stuttgart, Strecher.
WENINGER, J. 1927. Eine morphologisch-anthropologische Studie durchgeführt an 100 westafrikanischen Negern. Wien.
WENTWORTH, J. H., P. K. SMITH, AND W. U. GARDNER. 1940. *Endocrinology* **26**: 61.
WILDER, H. H., AND M. W. PFEIFFER. 1924. *Proc. Amer. Acad. Arts & Sci.* **59**: 443–603.
WOOD-JONES, F., AND T. D. CAMPBELL. 1924. *Trans. Roy. Soc. So. Australia* **48**: 303–312.
WÜRZINGER, S. 1928. *Zeit. mensch. Vererb. KonstLehre* **13**: 713–778.

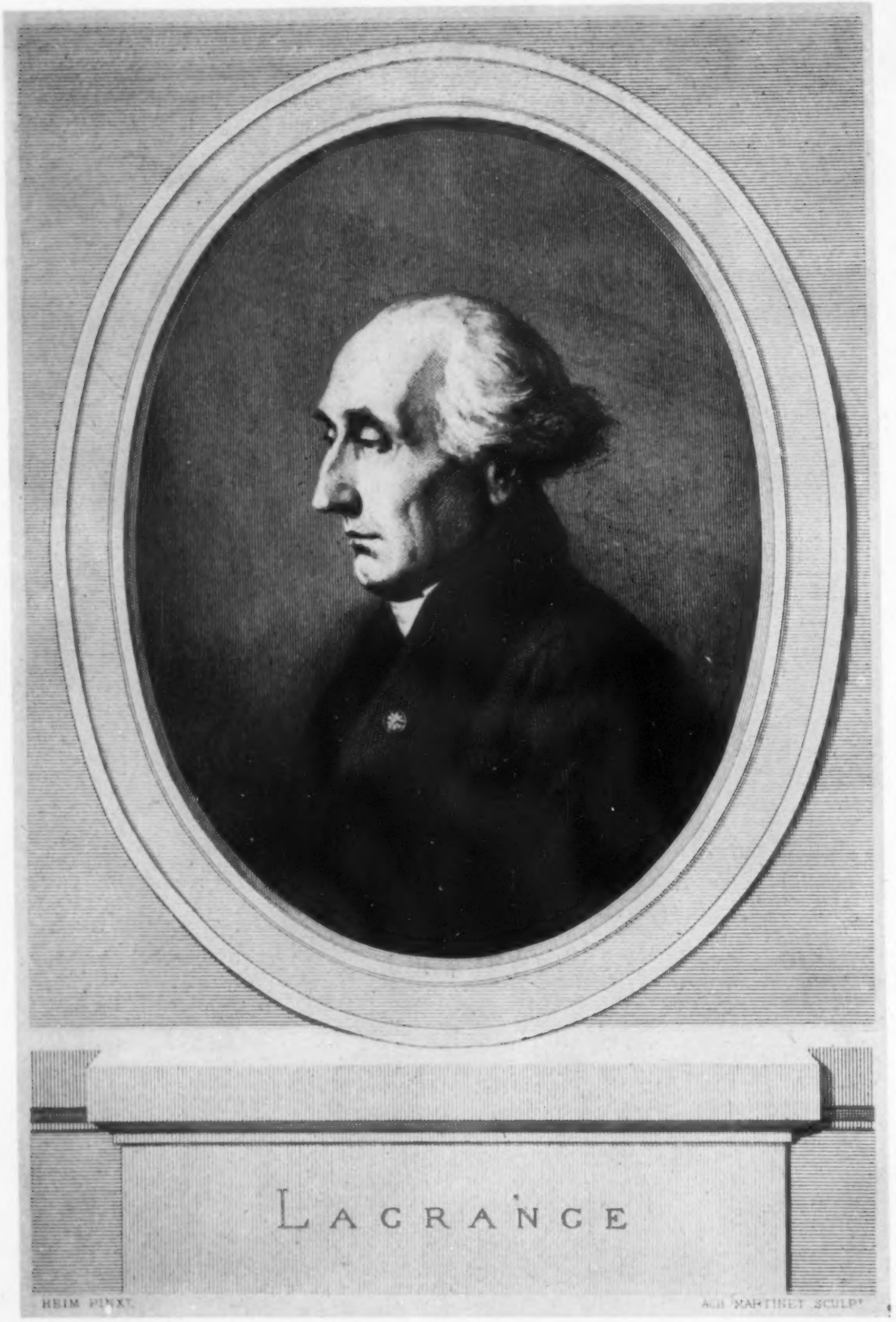


FIG. 1.

LAGRANGE'S PERSONALITY (1736–1813)

GEORGE SARTON

Professor, Harvard University

CONTENTS

I. Introduction 457
II. Biographical summary 457
III. Lagrange's nationality 459
IV. His use of the French language 460
V. His scientific work 461
VI. First fruits 461
VII. His modesty 464
VIII. Importance of academies of science in the eighteenth century 466
IX. Lagrange's life in Berlin 467
X. Mécanique analytique 469
XI. His two marriages 475
XII. His friendships 476
XIII. Various traits 479
XIV. His religion 480
XV. His interest in science and in the history of science 481
XVI. His views on mathematics 482
XVII. He weathers the French Revolution. His last activities 483
XVIII. His temperament, health, and death 488
APPENDICES
XIX. Collected works 490
XX. Manuscripts 491
XXI. Correspondence 491
XXII. Letters and other autographs in the United States 492
XXIII. Biographical sources 494
XXIV. Iconography 496

I. INTRODUCTION

WHEN the memory of scientific achievements has been enshrined by posterity, it is natural that one should be curious about their creator, even in a very personal way, and that one should ask oneself what kind of a man he was. Such curiosity is not always easy to satisfy, least of all in the case of a "pure" mathematician such as Lagrange.

How can we know at all what he was like? The sources of information are the contemporary documents and biographies, the references to him made by people who had actually known him, or had known persons close to him, autobiographical notes, remarks inserted in scientific memoirs (generally in the prefaces) and letters. Every detail not derived from such sources, which later writers have added to embellish or spice up their narratives, should be disregarded, for surely we want to know the real facts not the fancies of other people.

In the case of men who enjoyed sufficient popularity, there circulate anecdotes which deserve to be collected even if they be apocryphal. Those anecdotes do not help us to know the man, but they enable us to realize what his contemporaries thought of him, and to measure his influence.

Critical readers may wish to examine my credentials before proceeding further; they will find at the end of this paper a bibliography and a list of the published and unpublished texts which were used by me as far as was possible in war time. After having consulted those lists they will be able to decide whether it is worth their while to read this paper or not.

II. BIOGRAPHICAL SUMMARY

Joseph Louis Lagrange[1] (-Tournier) was born in Turin, Piedmont, on January 25, 1736 (XIV: 286).[2] He was baptized in the parish church of Sant Eusebio five days later; in the certificate of baptism, written in Italian, he is called Lagrangia Giuseppe Lodovico. The form Lagrangia was an arbitrary translation of a French name. One of his ancestors originated from Touraine and was in the service of Louis XIV, whom he abandoned to attach himself to Charles Emanuel II, Duke of Savoy from 1638 to 1675. From that time on (before 1675) the family had been established in Piedmont and the descendants of the French ancestor had married Italian women. The Duke of Savoy had created for the grandfather a new office, the treasury of construction and fortifications, and that office remained in the Lagrange family until its abolishment (1800).

[1] The name is also written de la Grange, de Lagrange, La Grange (see our figs.). The particles *de, de la*, suggested nobility; sometimes that suggestion was useful, sometimes it was dangerous. All such names changed frequently in France during the political vicissitudes of the eighteenth and nineteenth centuries. Moreover, proper names were not as fixed as they are now; they might be translated or modified in various ways. It is a waste of time to bother about those variations. In this article he is always called Lagrange. His main Christian name was Louis; in Italian Lodovico, Ludovico, or Luigi.

[2] In a symbol (*X*: *y*) the *X* refers to the volume number and the *y* to the page number of the *Oeuvres de Lagrange*.

Lagrange's father, Giuseppe Francesco Lodovico Lagrange, had married Maria Teresa Gros, daughter of a rich physician of Cambiano, a village near Turin. They had eleven children of whom only two survived, the eldest, Giuseppe Lodovico, and the youngest.

The family was well-to-do but suffered heavy financial losses during Lagrange's boyhood. The latter was educated in the Athenaeum of Turin,[3] and his first teachers of physics and mathematics were Giovanni Battista Beccaria and Filippo Antonio Rivelli. It is said that it was a paper written by Edmund Halley[4] which awakened his mathematical vocation; be that as it may, that reading was probably like the infinitesimal shock which causes a saturated solution to crystallize; may be also that it was the first original paper (vs. textbooks) to fall in the boy's hands. His mathematical interest was soon intense and exclusive. We may assume that his teachers acquainted him with the *Instituzione analitiche ad uso della gioventù italiana* (2 v., Milano, 1748) of Maria Gaetana Agnesi (1718–1799)[5] which provided then the best approach to analysis. This enabled him to master with incredible speed the works of Newton, Leibniz, Euler, the Bernoulli, d'Alembert. At the age of eighteen he was ready to fly with his own wings, and he began his mathematical correspondence with Fagnano and with Euler. We shall come back to that later on. In 1755 he was appointed a teacher in the Royal School of mathematics and artillery in Turin. In 1759 he helped to found the academy of Turin, and began his long friendship with d'Alembert. He had then already established himself—at the age of twenty-three—as the most promising younger mathematician of his time.

In 1763 the Marchese Domenico Caràccioli, having been appointed ambassador of the King of Naples to the King of England, left Italy to take office in London, and invited Lagrange to travel with him.[6] They arrived together in Paris, where Lagrange made the personal acquaintance of d'Alembert, Fontaine, Clairaut, and others; a serious illness prevented him from continuing to London with the ambassador. We know of his stay in Paris and of the (undefined) illness through a noble letter written by d'Alembert to Madame Geoffrin,[7] bespeaking her kindness for the young Italian. This letter does so much honor to d'Alembert and Lagrange that it is worth while to translate it.[8]

Saturday morning, 7 Jan. 1764

You can do, Madam, a meritorious deed and worthy of you for a man of the highest merit. Mr. de la Grange a young geometer of Turin, has been here for six weeks. He was taken ill, rather seriously, and needs, not financial help, for the marquis Caraccioli has given orders before leaving for England to make sure that he [Lagrange] would be in want of nothing, but a few marks of interest from the part of his own country. He is a man of the rarest talent in Geometry, far above anything Italy has in that line, and at least the equal of all the best in the rest of Europe. Mr. Fontaine, my colleague at the Academy of Sciences, who as you know is chary of praise, will give you the same testimony, Mr. Clairaut and all those who are familiar with the works of Mr. de La Grange will certainly tell you the same thing.

It is a treasure which Turin owns, without perhaps realizing its value. You are acquainted with the Ambassador of Sardinia, and you should entreat him to give Mr. de la Grange in his present situation a token of interest and to write to the Sardinian court all the good that one thinks of him here, and all the esteem which the greatest Geometers have for his talent and his person. Mr. Ambassador will do thus a deed worthy of himself; he will oblige a man of the highest merit, who needs most in the present circumstances consolation and encouragement. If I had the honor of knowing the Ambassador of Sardinia, or of meeting him, I would not hesitate to confirm my-

[3] The certificate of matriculation dated 1750 is reproduced in facsimile in *Isis* **28**: 368, 1938.

[4] An instance of the excellence of the modern algebra in the resolution of the problem of finding the foci of optick glasses universally (*Philos. Trans.* **17** (205): 960–969, Nov. 1693).

[5] There is no reason to believe that Lagrange knew that remarkable woman personally, but he refers to her great book in a letter to Fagnano dated October 30, 1754 (Fagnano, 1912: 203). She was only thirty when her famous *Instituzione* appeared, but a few years later (*ca.* 1752) she withdrew from the world and devoted herself entirely to religion and charity. In 1762, the academy of Turin tried to bring her back to mathematics but did not succeed. Her mathematical life lasted a little more than ten years, her religious life almost half a century.

[6] The passport which Lagrange obtained from Charles Emanuel, Duke of Savoy and King of Sardinia from 1730 to 1773, has been preserved. See facsimile of it in *Isis* **28**: 369, 1938. It is written in French and dated Turin, October 29, 1763.

[7] Marie Thérèse Geoffrin, née Rodet (1699–1777), conducted one of the most famous salons in Paris and was very influential. She had paid the debts of prince Stanislaw Poniatowski and thus saved him from prison; he was later the last king of independent Poland, Stanislaw II (1764–1795). Madame Geoffrin went to Warsaw in 1766 to see him, though she was then an elderly woman.

[8] It was first published by Charles Henry in Boncompagni's *Bullettino* **19**: 130, Rome, 1886. It was reprinted by Guareschi (1913: 13). It is not reprinted in the *Oeuvres*.

self what I have the honor of writing to you, being sure of finding him in the mood to favor and encourage merit, especially when it is carried to such a high degree as it is in Mr. de la Grange. I, as well as Geometry and those who cultivate it, we would all have a true obligation to you if you would take this matter at heart, and I will remain all my life with as much gratitude as respect, Madam,

Your very humble and obedient servant

D'ALEMBERT

The letter is as poorly written as it is strongly felt. It illustrates not only Lagrange's early fame but also his curious relationship to the authorities of his own country. A strange situation indeed. Lagrange was protected by the ambassador of Naples (a different country then from his own), and he needed the circuitous efforts of d'Alembert and Madame Geoffrin to obtain some mark of sympathy from the official representative of his own country. It is possible that if we had more documents of this kind we would understand better Lagrange's coldness toward Piedmont and Savoy.

We don't know much more about his first stay in Paris, but it is clear that during those many weeks his friendship with d'Alembert had the time to ripen. He returned to Turin via Geneva, where he stopped at the Délices to dine with Voltaire, and reached his home in May 1764 (XIII: 10).

It is probable that he felt rather lonely in Turin after his return, and he must have found the teaching of mathematics at the military school irksome; he was certainly anxious to obtain a position somewhere else. In 1766 upon the joint recommendation of Euler and d'Alembert, Frederick II invited him to come to Berlin. He accepted readily and left on August 21, (XIII: 77) for Paris, where he remained only a couple of weeks and renewed his friendship with d'Alembert. He sailed to England to pay a brief visit to the Marquis Caràccioli. He arrived in London on September 20 (XIII: 77) then sailed to Hamburg and reached Berlin in the beginning of November 1766. He did not realize then that he would never return to his native city. He remained in Berlin until 1787, when he was elected to a similar position in the French Academy of Sciences. He spent the rest of his life in Paris, where he died on April 10, 1813.

World-shaking events happened during his life time, but he quietly continued his chosen task almost until the end. Out of seventy-seven years, he had spent the first thirty in Turin, then twenty-one in Berlin and twenty-six in Paris.[9]

III. LAGRANGE'S NATIONALITY

There has always been much intellectual jealousy between the French and the Italians, and Lagrange has always been one of their bones of contention. Should we call him an Italian or a Frenchman?

In a way, the last sentence of the previous chapter has already answered the question, as well as introduced a third claim. Lagrange was born in Italy of Italian parents (the reference to his French great-grandfather is hardly relevant); he was educated in Italy and there began his magnificent career. He is in the fullest sense a son of Italy. The thirty years he spent in Turin were years of intense preparation for his chosen task. His best work, however, including the writing of the *Mécanique analytique*, was done in Berlin. When he arrived in Paris in 1787, Queen Marie Antoinette was especially kind to him, because he had been recommended to her as a German. Did he not come straight from Berlin, where he had lived twenty-one years? He worked less in Paris, or the work which he did there was less profound, but he reaped his harvest. The Parisian years included the revolutionary nightmare, but when the storm was over many honors were heaped upon his head by the First Consul and the Emperor.

Lagrange never returned to his native country, but he remained in touch with Italian men of science. In 1781, when Colonel Antonio Lorgna invited him to help establish the Società italiana delle scienze (Società dei XL), Lagrange answered (XIV: 263) that he was anxious to merit that honor and to prove himself a good fellow citizen (*buon compatriota*). However, he refused various positions that were offered to him in Naples or Turin. Obviously, he preferred to be in Berlin or in Paris than in his own country.

Thanks to d'Alembert he had been elected to membership of the Académie des Sciences as early as 1772, being then given the very high distinction of "associé étranger" at the age of thirty-

[9] Contrast with Euler, who spent his first twenty years in Basel, and whose sojourn of twenty-five years in Berlin was sandwiched between two stays in St. Petersburg of fourteen and seventeen years; the total being almost the same as Lagrange's, seventy-six years. As to d'Alembert he remained during the whole of his somewhat shorter life (sixty-six years) in Paris.

six[10] (XIII: 240). This meant that the French considered him a foreigner, albeit a very distinguished one. He was "associé étranger" for fifteen years, until his arrival in France in 1787 when he was given the title of "pensionnaire vétéran" and the right to vote in every deliberation (I: xxxvii). At the time of the Revolution the National Assembly (1789) continued to pay him the pension which the king had granted (XIV: 284). When the Convention decreed in 1793 that resident foreigners born in enemy countries should be arrested Lavoisier applied to Lakanal in Lagrange's favor (September 7, 1793; XIV: 314) and the latter was left undisturbed. This fact is quoted here to prove that from the French legalistic point of view Lagrange was then a foreigner, an "enemy alien."

In Lavoisier's letter it was stated that Lagrange had chosen France as his adopted country (*sa patrie adoptive*) and that was true. When the Consulate was established (Constitution of the year VIII, December 24, 1799) Lagrange was one of the senators. From that time on he must be counted as a Frenchman. Later, Napoleon gave him the dignity of an imperial count. Before criticizing him for accepting such favors, one must bear in mind that it would have been impossible for him to continue his residence in Paris without accepting them and that his own country had ceased to exist as an independent unit, Piedmont having been annexed to France on 26 frimaire an VII (= December 16, 1798). At the time of the annexation his father was still living in Turin, being then ninety years of age. Talleyrand, who was then Minister of Foreign Affairs of the Directory, ordered Mr. d'Eymar, commissary for Piedmont, to send an official delegation to the old man. The commissary went with many generals and other distinguished men and congratulated the elder Lagrange for having given birth to such an illustrious son. A few years later another Frenchman, a member of the Institut, visited him again and as he was praising the son, the old man answered "Yes my son is great before men, may he be great also before God!" He died shortly afterwards (*ca.* 1803) at the age of ninety-five.[11] Piedmont was divided by the French invaders into six departments, Turin being the main city of the Département du Pô; that situation obtained until 1814, that is, until after Lagrange's death.

[10] He replaced the Italian physician, Giovanni Battista Morgagni (1682–1771).

[11] For further details see Maurice (1819: 165).

To sum up we may say that Lagrange was an Italian, who became a Frenchman toward the end of his life. The whole world owes his genius to the country where he was born and educated, but two other countries are entitled to share his glory.

One cannot say that Lagrange ever was a patriot either French or Italian. He respected the laws and did his duty wherever he was, but he was more interested in mathematics than in France or Italy. His insensitiveness to his family and native country is sufficiently proved by the fact that during thirty-seven years after his departure from Turin he never found time to pay a visit to his father, and in forty-seven years he never found time to return to his native land and see the comrades and colleagues of his youth. He had enough money to travel and was largely the master of his time. It must be added, however, that during the revolutionary and war years travel would have been difficult if not impossible and that when peace was restored he was probably embarrassed by conflicting loyalties.

IV. HIS USE OF THE FRENCH LANGUAGE

It has often been argued that Lagrange's almost exclusive use of the French language, proves that he was a Frenchman at heart. The fact is correct but the conclusion is wrong. It is true that with the exception of his first paper (the *Lettera* of 1754) every scientific paper of his was composed in French; the vast majority of his letters were written in the same language, very few in Italian, fewer still in Latin (his first letters to Euler), none in German. He wrote in French even to Italians established in Italy. I have read two letters of his written in French to his own brother who was then practising law in Turin.[12]

Lagrange was not alone in preferring French to his native language. The very academy which he helped to found in Turin was attaching more importance to French than to Italian and even than to Latin. The first volume of its *Miscellanea* (1759) had a Latin title page, but its dedication to the Duke of Savoy was in French; from the second volume on (1761) the title page was printed in French,[13] but the bastard title continued in Latin. The situation was similar in Germany. No one would suggest that Frederick

[12] The one in *Oeuvres* (XIV: 303), the other published by Guareschi (1913).

[13] *Mélanges de philosophie et de mathématique de la Société royale de Turin.* Note that the private society had now become a royal one.

the Great was not a good Prussian, yet he placed the French language far above the German. In fact, he was the outstanding promoter of French letters abroad. He decided that French should be the official language of the Prussian Academy; in the school[14] which he had founded for the training of officers French was the language of instruction and students were at first forbidden to speak German. Linguistic nationalism did not yet exist, and French had become the polite language of continental Europe. To put it otherwise the use of French implied a social rather than a national distinction; gentlemen and scholars spoke French and thus emphasized their superiority over the canaille.

The establishment of French as the international language of polite society was a victory not only over the other vernaculars, but also over Latin. Humanists may be scandalized to be told that the gradual replacement of Latin by French was a triumph for the humanities, but so it was. Latin had remained so artificial that few people (most of these clerics) were able to use it with any spontaneity and native grace. The average layman obliged to write in Latin was cramped and maladroit, and if he tried to be elegant it was even worse, for he could not help falling back on alembicated sentences and moth-eaten ornaments; to express the lights and shades of his thought or his subtlest feelings was generally out of the question. Under those conditions there was no possibility of individual style and the humanities became a fraud. This is completely proved by the absence of any kind of Latin literature after the Renaissance in spite of the existence of many thousands of volumes written in Latin, some of them by men of great distinction. Not a single one of those books has remained a living part of the world literature.[15] No language can be really alive which is not spoken by women. Latin was not used by women, while French was; therefore, Latin was bound to be defeated.

As a matter of fact Lagrange was francophile, but his use of the French language does not prove it. One could very well express anti-French feelings in French, even as today the English language is not unfrequently used to vilipend England or America.

Lagrange himself has not declared his views on the matter of language, but they were probably very similar to those which his friend d'Alembert explained so well in his famous essay on the "latinity of moderns."[16]

A final remark. We should bear in mind that Lagrange came from Piedmont and that the dialect of that "paese" has resisted linguistic unification more than any other northern dialect. To be sure Italian was taught in the schools as thoroughly as Latin and French, but the upper classes spoke French or Piemontese rather than Italian. That situation continued almost until the end of last century. For example, when Cavour and his king confabulated we may be sure that they used their homely dialect rather than the national language. That dialect by the way is in some respects closer to Provençal (or Catalan) than it is to other Italian dialects.

V. HIS SCIENTIFIC WORK

No attempt will be made here to discuss his scientific work in its entirety, for it could not be done without extending this essay considerably. It will suffice to remark that it dominated mathematical thought for forty years, say from 1759 to the end of the century. It is not possible to explain the history of mathematics during that period without referring to Lagrange repeatedly, and vice versa it is impossible to give a good account of Lagrange's achievements without bringing in his contemporary rivals. The fourth volume of Moritz Cantor's *Vorlesungen* is devoted exclusively to what might be called the Lagrangian period (1759–1799). The mathematicians most often quoted in that enormous volume are Euler, Lagrange, d'Alembert, and Laplace (in that order), but Lagrange is the only one of those four giants who covers the whole period, for Euler and d'Alembert died in 1783 and Laplace entered the lists only in 1774.

VI. FIRST FRUITS

Lagrange offered the handsel of his genius to the mathematician, Giulio Carlo da Fagnano (1682–1766)[17] in the form of a letter dated Turin,

[14] I do not call it academy to avoid confusion, but its usual name was Académie civile et militaire des jeunes gentilhommes.

[15] Purely scientific books like Newton's *Principia* are of course excluded.

[16] Sur l'harmonie des langues, et en particulier sur celle qu'on croit sentir dans les langues mortes et à cette occasion sur la latinité des modernes, *Mélanges de littérature, d'histoire et de philosophie* V: 525–568, Amsterdam, 1767. That essay may be read also in the *Oeuvres de d'Alembert*, III, Paris, 1805.

[17] Volume III of the works of Fagnano includes the correspondence, notably with Lagrange (1754–1759), and his

July 23, 1754, and printed separately in that city the same year.[18] That "letter" was written in Italian and is the only scientific paper of Lagrange written in that language. As set forth in its very title it explained "a new series for the differentials and integrals of every grade corresponding to the Newtonian series for powers and roots." It was a very nice discovery for a young man of eighteen to make, but unfortunately it had been made by Leibniz almost half a century before.[19] If we may believe Arago, when young Lagrange discovered that he had been anticipated by Leibniz, he lost consciousness and almost gave up his mathematical studies.[20] This is very probably an exaggeration. To be anticipated by Leibniz was a misfortune and a complimentary one at that, rather than a crime. There is no indication that Lagrange interrupted his mathematical studies; he did not even abandon that particular topic and came back to it in greater style as late as 1772.[21]

To return to the first fruits, Lagrange was so little discouraged that a year after the publication of the *Lettera* we find him writing to Euler, and communicating to him what might be called the first idea of a new branch of analysis, the calculus of variations.[22] According to his own account he had been meditating Euler's book on isoperimetric problems[23] and had thus been led to his momentous discovery. Euler, who we must remember was almost thirty years older than the unknown Italian youth, greeted the latter's discovery with genuine enthusiasm. Wrote he (Berlin, September 6, 1755; see XIV: 144) "Perlectis tuis postremis litteris, quibus Theoriam maximorum ac minimorum ad summum fere perfectionis fastigium erexisse videris, eximiam ingenui tui sagacitatem satis admirari non possum. . . ." In short, "I cannot admire you as much as you deserve!" It is easy to imagine what such a testimony coming from the illus-

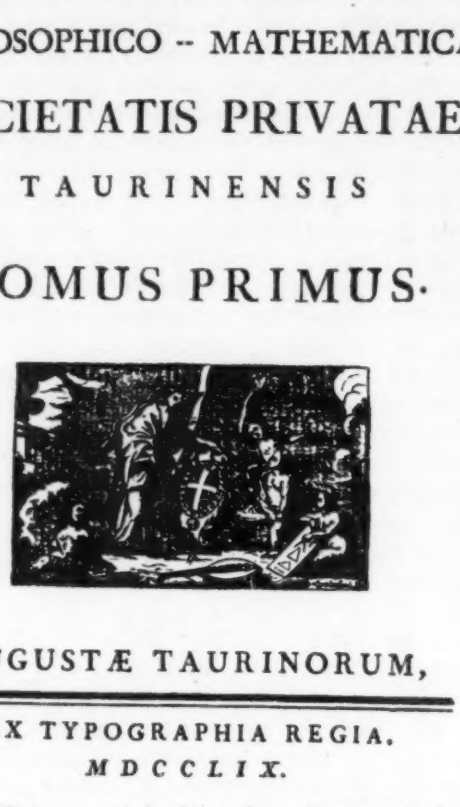

MISCELLANEA
PHILOSOPHICO -- MATHEMATICA
SOCIETATIS PRIVATAE
TAURINENSIS
TOMUS PRIMUS.
AUGUSTÆ TAURINORUM,
EX TYPOGRAPHIA REGIA.
M D C C L I X.

FIG. 2. Title page of the first volume (in quarto) of the *Miscellanea Taurinensia*, Turin, 1759.

biography. Fagnano initiated the theory of elliptic functions and found the pretty formula $\pi = 2i \log \frac{1-i}{1+i}$, in both cases showing the way to Euler.

[18] Little pamphlet of eight pages small quarto, printed by the Royal Press, Turin 1754. See description of it by Prince Boncompagni in his *Bullettino* **1**: 183, 1868. Boncompagni knew of five copies of it, one of them in his own possession; it is probable that only very few copies were printed. In spite of various entreaties I have not been able to locate a single one in this country. The *Lettera* was reprinted in the *Storia letteraria d'Italia* (for the second half of 1754, **10**: 107–112, Modena, 1757). It is reprinted also in the *Oeuvres* (VII: 583–88).

[19] Symbolismus memorabilis calculi algebraici et infinitesimalis in comparatione potentiarum et differentiarum etc., *Miscellanea Berolinensia, Berlin Akad-Wiss.* **1**, 1710.

[20] Arago's account in his life of Fresnel, 1830 (*Oeuvres d'Arago* I: 119, 1854) contains other incorrections. Arago was an incorrigible romanticist and a careless historian; he wrote very easily and, like the popular writers of today, was more eager to polish a sentence or make a witty remark than to find and tell the truth.

[21] Sur une nouvelle espèce de calcul relatif à la différentiation et à l'intégration des quantités variables, *Nouveaux Mémoires de l'Académie de Berlin*, 1772, reprinted in *Oeuvres* III: 441–476. It is typical of Lagrange's reticence that he does not mention at all his juvenile paper of 1754. Placido Tardy, Intorno ad una formola del Leibniz, *Bull. di bibliografia e di storia delle scienze matematiche* I, 177–186, Roma, 1868.

[22] That was his second letter to Euler; it is written in Latin and dated August 12, 1755 (XIV: 138–144).

[23] *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, sive Solutio problematis isoperimetrici latissimo sensu accepti*, Lausanne, 1744. Facsimile of title page in *Isis* **20**: 72, 1933.

A SON ALTESSE ROYALE
MONSEIGNEUR LE DUC DE SAVOYE.

MONSEIGNEUR

L'Ouvrage que nous avons l'honneur de vous préſenter eſt le prémier fruit des tra-

FIG. 3. First page of the dedication of the *Miscellanea*, vol. 1, 1759.

trious master must have meant to the young man; yet in the letter which he wrote shortly afterwards to Fagnano, he refers to it very simply (*onorevolissima lettera*) without repeating Euler's extraordinary praise.[24] Nor was this empty praise, for a year later Euler obtained Lagrange's appointment as foreign member of the Berlin Academy and sent the glad tidings immediately to him (Berlin, September 2, 1756; XIV: 156).

On October 2, 1759, Euler wrote again reiterating his admiration of the new method but adding "The importance of that subject has stimulated me to develop, aided by your lights, an analytical solution which I will keep secret as long as your own meditations are not published, lest I take away from you a part of the glory which you deserve."[25]

[24] Lagrange to Fagnano, December 24, 1755. Fagnano, 1912: 205–207.

[25] Rei dignitas me excitavit ut tuis luminibus adjutus, ipse solutionem analyticam conscripserim quam tamen celare statui, donec ipse tuas meditationes publici juris feceris, ne ullam partem gloriae tibi debitae praeripiam (XIV: 163).

AL CHIARISSIMO

SIG. LUIGI
DE LA GRANGE
TORINESE

Direttore della Claſſe Matematica nella Reale Accademia delle Scienze e Belle Lettere di Berlino

EPISTOLA

Di un ſuo Concittadino, ed Amico.

IN BOLOGNA

Nella Stamperìa di Lelio dalla Volpe. 1767.
Con licenza de' Superiori.

FIG. 4. Title page of Italian poem written by Giuseppe Maria Boccardi in Lagrange's honor.

In the meanwhile, Lagrange was better known also among the men of science of Turin, and together with the chemist Saluzzo de Monesiglio and the anatomist Cigna he was preparing the organization of the private scientific society, which was to become later the Turin Academy of Sciences.[26] The first volume of the *Miscellanea Taurinensia* (1759) contained no less than three papers of Lagrange. The most important of those papers was one entitled "Researches on the nature and propagation of sound" (II: 39–148)

[26] Il conte Angelo Saluzzo di Monesiglio (1734–1810), Gian Francesco Cigna (1734–1790). Biographies of both and history of the Turin Academy will be found in the centennial publication, *Il primo secolo della R. Accademia delle scienze*, 1783–1883.

wherein he discussed in a new way the problem of vibrating strings which had already been attacked by Euler, d'Alembert and Daniel Bernoulli. Copies of that volume were mailed eagerly to Euler and d'Alembert (end of July 1759). As soon as he received it Euler wrote (October 23, 1759), this time in French, discussing Lagrange's treatment of that problem. D'Alembert's acknowledgment of the same volume was equally cordial; he ended his letter of thanks dated Paris September 27, 1759, with the gracious words "Adieu, Monsieur, vous êtes destiné, si je ne me trompe, à jouer un grand rôle dans les sciences, et j'applaudis d'avance à vos succès. . . ." (XIII: 4).

Lagrange was only nineteen when he invented the calculus of variations (that is, the basic idea of it), and as that calculus and himself matured he was conscious of that achievement.[27] On November 20, 1769—that is, when he was thirty-three—he wrote to d'Alembert, "I always regard that method as the best I have ever done in Geometry" (XIII: 154).

Let me quote a few more of the early works of Lagrange, to wit his *Théorie de la libration de la lune* (1764; VI: 5–61) for which he obtained a prize from the Parisian Academy. This contains the first idea of the principle of virtual velocities which was to be one of the fundamental principles of his *Mécanique analytique* (1788: 8–12). In 1765, he took part in another competition organized by the same academy and his *Recherches sur les inégalités des satellites de Jupiter* (1766; VI: 63–225) caused him to be proclaimed the laureate for the second time.

In the meanwhile he had published his first paper in the *Mémoires de l'Académie de Berlin*, on tautochronous curves (1765; **11**: 317–332). Thus in 1766—at the age of thirty—Lagrange was already known as a mathematician of the first order all over Europe, for the papers he had published in the collections of Turin, Paris, and Berlin received the most complete publicity among the relatively few scientists who were interested in them. Moreover, he had won the allegiance of the two accepted masters of the day, Euler and d'Alembert. He was anxious to leave Turin and find a better position. It was at this time that he left his native country never to return.

[27] His first printed explanation of it appeared only in the second volume of the *Miscellanea Taurinensia, R. Accad. Sci. Torino*, Essai d'une nouvelle méthode pour déterminer les maxima et les minima des formules intégrales indéfinies (II, 1760–1761; *Oeuvres* I: 335–362).

Maupertuis who had been President of the Berlin Academy since 1746 had died in 1759 and had not been replaced. Euler, Director of the Mathematical Section, had discharged the presidential duties even before Maupertuis' death, but Frederick the Great who knew his foibles would not have him as the latter's successor. It is probable that this caused or increased Euler's dissatisfaction, and his desire to place himself again under the wings of Ekaterina II in St. Petersburg. Frederick would have liked to appoint d'Alembert, but d'Alembert did not wish to leave Paris. Both Euler who did not want to stay and d'Alembert who did not want to go recommended Lagrange for the post, not of President, but of Director of the Mathematical Section. D'Alembert's recommendation was especially powerful, for he was the King's trusted friend. Harnack goes so far as to say that d'Alembert was the secret president of the Berlin Academy for twenty years (1763–1783).[28] On October 2, 1766, Lagrange was appointed a regular member of the Prussian Academy, and soon afterwards Director of the Mathematical Section (that is, he was Euler's, not Maupertuis', successor; Maupertuis was never formally replaced).

This was a great triumph for a Piemontese mathematician of thirty years! His countrymen fully realized it, witness the Italian poem which was published in Bologna in 1767. It is a pamphlet of thirty-one pages, the title page of which is here reproduced. It was published anonymously but with a preface by Francesco Maria Zanotti (1692–1777), physicist and man of letters, President of the Institute of Bologna; the unnamed rhymester was Giuseppe Maria Boccardi.

Lagrange arrived in Berlin on October 27, 1766, he was officially received and installed on November 6 and read a short compliment in French (XIV: 316). On November 12, he presented his first scientific paper to the Academy, a memoir on the transit of Venus of June 3, 1769 (II: 335–374).[29] The next twenty-one years of his life were spent in Berlin.

VII. HIS MODESTY

When Lagrange arrived in Berlin, he was already illustrious in spite of his youth, and what is more to the point he had originated seminal ideas

[28] Harnack I: 466, 1900.

[29] For further details on that transit see Sarton, Vindication of father Hell, *Isis* **35**: 97–105, 9 fig., 1944, chiefly pp. 99–102.

in various fields and prefigured a large part of his creative life. While he pursues the even tenor of his work we may ask ourselves a few questions concerning his character.

The main traits, noticed by every one who came in contact with him, were his modesty and sweetness of temper. In a characteristic letter dated June 30, 1772, Frederick the Great wrote to d'Alembert

> Vous distribuez des billets de grand homme à ceux qui se distinguent parmi les nations étrangères. Je suis bien aise que notre La Grange soit de ce nombre; je suis trop ignorant en Géométrie pour juger de son mérite scientifique, mais je suis assez éclairé pour rendre justice à son caractère plein de douceur et de modestie (XIII: 246).

The king had been able to compare two of his mathematicians who were in that respect as different as could be, Lagrange and the Alsatian, Johann Heinrich Lambert (1728–1777). Lambert's talent was great enough, his conceit, much greater.[30]

The modesty of the former and the vanity of the latter are explained in a similar way; Lagrange's modesty was the natural result of ceaseless self-criticism, while Lambert was apparently incapable of such criticism and swallowed himself whole.

Writing to Condorcet (1743–1794), then Perpetual Secretary of the Paris Academy, on November 23, 1773, to announce the presentation of a new memoir he remarks

> You will do me a real favor in examining my memoir severely and in suppressing it if you think it advisable to do so. You could then mail it back at the earliest opportunity, and I undertake to send you another on some other subject. Above all, I beg of you not to present it to the Academy before having obtained the advice of M. d'Alembert, to whose judgment I submit myself completely (XIV: 15).

He writes in the same vein to d'Alembert on October 1, 1774, begging him to suppress a memoir if he does not think it worthy of the attention of the Paris Academy.

> I beseech you to be convinced that this is not a mere compliment from my part, made with the view of eliciting the more praise for my Work that I seem to lay fewer claims; I flatter myself that you know my manner of thinking well enough to believe me incapable of such subtlety. I am duly sensible of your approval, which I even regard as the greatest and almost the unique reward of my efforts, but I am always so little satisfied with what I am doing and care so little for it that my great distrust of the merit of my Works is not astonishing (XIII: 291).

To young Laplace (1749–1827) on September 1, 1777:

> I have always considered Geometry as an object of amusement rather than of ambition and can assure you I enjoy the works of other people much more than my own, for I am always dissatisfied with the latter; hence you see that if you are free of jealousy because of your own success, I am not less so because of my character (XIV: 71).

To d'Alembert, on January 1, 1781:

> As I am in no hurry and work for my pleasure rather than as a matter of duty, I do like the great lords who are building: I do, undo and redo many times, until I be passably satisfied with my work, nevertheless that happens only on very rare occasions (XIII: 360).

To these extracts taken from the Berlin letters, let me add two more taken from letters written by Lagrange in Paris, to Father di Caluso.[31] In 1784 he writes

> I become every day harder to please with my papers and instead of increasing the number of the printed ones, I would like to be able to cancel some of them, or at least to reduce them to a few pages.

On April 28, 1788:

> Permit me to beg of you the acceptance of a copy of my *Mécanique analitique* which has just appeared. I would like to have something better to offer to you in order to convince you of my esteem and to deserve yours. . . .

This was perhaps going a little too far and calls to mind the conventional exaggeration which eighteenth-century courtesy required. However, Lagrange's modesty was profoundly genuine. We should remember that what the average man, who is neither too vain nor over modest, requires above all is appreciation, genuine not second-

[30] Lambert had been elected a member of the Berlin Academy a year before Lagrange, in 1765, but he died so young (at forty-nine) that their association lasted hardly more than a decade. Amusing stories illustrating Lambert's conceit are told by Thiébault, 1804 (V: 24–36).

[31] Il abate Tommaso Valperga di Caluso dei Conti di Masino (1737–1815), officer on the fleet of the Order of Malta, then Oratorian in Naples to 1768, later in Turin. Mathematician, philosopher, man of letters. Lagrange wrote to him in French, not in Italian.

hand appreciation. Alas! mathematicians can expect none except from their peers. It is probable that Lagrange was sick and tired of the praise lavished upon him by ignorants who did not even know what they were talking about.

One may object that it was easy enough for Lagrange to be modest as his genius was universally recognized and he had reached the top of the ladder early in life. He had no need whatsoever of asserting himself. Think of it, at thirty he was Director of the Mathematical Section of the Berlin Academy and six years later (1772) he was "associé étranger" of the Paris one (XIII: 240). What more could he want? Nevertheless, conceited people like Lambert or Laplace are never satisfied. Their conceit increases with their success if not faster; on the contrary it is characteristic of men as fundamentally modest as Lagrange that success excites their self-criticism and their self-distrust.

VIII. IMPORTANCE OF ACADEMIES OF SCIENCE IN THE EIGHTEENTH CENTURY

Lagrange's life was closely connected from first to last with academies. He was one of the creators of the Turin Academy and one of its active members from the beginning to 1785. He was a resident member of the Berlin Academy for twenty-one years and a collaborator to its memoirs from 1765 to 1803. He was a resident member of the Paris Academy from 1787 to his death and a contributor to its memoirs from 1764 to 1809. He was never outside the intimate circle of one of those academies and his active life might easily be divided into three periods concerning each of them in turnTurin1754–1766, Berlin 1766–1787, Paris 1787–1813.

In order to realize the meaning of this one must remember that academies were very different in the eighteenth century from what they are today. The Paris and Berlin academies especially were more like our research institutes than our modern academies, and they were not only the main research institutes but in those two countries almost[32] the only ones. The contemporary universities were scholastic, conservative, and timorous. The main scientific work of the eighteenth century was done by academicians rather than by professors.

[32] The word "almost" must be added because of the observatories. Some research of a more utilitarian nature was conducted also in hospitals and in military or naval establishments.

The recruitment of members was also very different in those eighteenth-century academies from what it is in the modern ones. The latter's problem is easy enough. They select scientists who have already become sufficiently famous, and that is not more difficult than to select the winning horses after the races have been run. On the contrary the old academies of Paris, Berlin, and we might add St. Petersburg were highly competitive institutions; it would not have helped them much to appoint illustrious old men who might have let them down, they must obtain the co-operation of the most promising scientists in their prime. That is why Euler was appointed a member of the St. Petersburg Academy at twenty, and Lagrange of the Berlin one at thirty.

The Prussian Academy had been founded by Leibniz in 1700, but had not been really organized until 1711; during the reigns of Frederick I (1701–1713) and Frederick William I (1713–1740) it had remained sleepy and sluggish. It is one of the merits of Frederick II, who ruled from 1740 to his death in 1786, that he vitalized the academy and made of it one of the outstanding research institutions in the world. He wanted *his* Academy to be one of the best, and he truly succeeded. He deserves to be called Frederick the Great because of his patronage of literature, art, and science. The history of his academy can be followed in detail in the excellent account which Adolf Harnack prepared for its second centenary. Interesting anecdotes concerning it are told by Thiébault and appear in the biographies of the academicians. Frederick II was in many respects a great man, but he was a dictator and a martinet, and he intended to control *his* academy as he controlled *his* regiments, and to keep a tight hand over *his* academicians as he did over *his* officers. In particular, elections were made at his bidding, and it was impossible to be elected without his favor (XIII: 166, 258). Academic and court intrigues were all mixed up. Lagrange had the good fortune of being d'Alembert's friend, d'Alembert being the King's most trusted adviser; though he was not in any sense an intriguer and was sufficiently protected if not by his merit, which few could appreciate, at least by his fame which was beyond dispute, still he found it useful to be remembered from time to time to the king in d'Alembert's letters (XIII: 165, 240); d'Alembert understood that perfectly and did his best to uphold his friend in the royal eyes. Euler's departure from Berlin in 1766 was

partly caused by his growing impatience with the royal whims and Frederick was glad to see him go. Frederick wrote to d'Alembert on July 26, 1766,

> I am thankful to you for having replaced in my Academy a one-eyed geometer by a two-eyed one.[33] . . . The modesty with which you compare yourself with sieur de la Grange enhances your merit instead of lowering it, and will not cause me to change my views and my esteem for you (XIII: 76).

IX. LAGRANGE'S LIFE IN BERLIN

A good picture of Lagrange as he was in Berlin has been left to us in the reminiscences of Dieudonné Thiébault. This Thiébault, born in 1733, was a Jesuit father until 1761 or 1762 when the order was expelled from France; in 1765 he received the chair of general grammar in the military academy created by Frederick II and became one of the King's familiars; he remained in Berlin until 1784 then returned to France; he died at Versailles in 1807. He wrote many books the most valuable being entitled *Mes souvenirs de vingt ans de séjour à Berlin*.[34]

The following passage is taken from that book (V: 39–42) or rather from the English version (II: 305–307, 1806). The English translation is free and somewhat abbreviated but correct as to essentials.

> At Berlin he married a lady who was his relation, and who possessed an excellent understanding and extraordinary sweetness of temper. His domestick life was tranquil and retired, as they both preferred the gratifications arising from the society of a small but select circle of friends. To the inestimable advantage of a peaceful home, M. de la Grange joined that of pursuing his publick career with the same serene tranquility. He was on no occasion accessible to any kind of intrigue or party spirit, and if some slight dissension took place in the academy, he never was concerned in it, nor did he seem acquainted with its existence. Nor should this trait be ascribed to timidity. I perfectly remember that M. de Schxxx, the minister, who was a man of excellent understanding, but of a passionate and haughty temper, having prevailed on the king to adopt the project of a fund for widows and M. de la Grange, having read a memoir to the academy in which he demonstrated that this fund would necessarily and speedily end in a bankruptcy, the minister sent a message to the former, importing, that instead of publishing his memoir he should have communicated it to government; to which the academician replied, 1st, that he had not published his memoir, but had contented himself with exercising a duty incumbent on friendship, by warning his colleagues against the danger of involving their interests in the above project; and 2dly, that not having engaged himself in the academy for the purpose of receiving the commands of ministers, he conceived he had nothing to do with waiting in their anti-chambers to offer them advice they had not required him to give; that it was their own business to look out for persons capable of furnishing principles for the calculations they stood in need of; that, in short, no blame could possibly attach to him, since he had not been consulted on the question. This reply, at once resolute and equitable, reduced M. Schxxx to silence.
>
> M. de la Grange observed a regular and uniform plan for the occupation of every day. His mornings were consecrated to reading and writing letters; immediately after dinner he devoted a few hours to different visits, or to his accustomed walk, which he took alone that he might pursue the pace he judged salutary for his health. At six in the evening he returned to his cabinet, where he shut himself in, that he might be sure of remaining undisturbed till midnight, when he took several dishes of tea before he went to bed.
>
> In these six hours of profound solitude it was that he performed the immense labours with which the memoirs of the academy are stored, and which have gained him a reputation so illustrious. What more shall I say? A philosopher remarkable for the equanimity of his conduct, ever wise, ever tolerant, blending with his genius for mathematicks a knowledge no less extensive than various in the different branches of literature, and with these the most genuine simplicity and gentleness of manners, he could not fail to be beloved and respected by all who knew him, and deeply lamented by those from whom he was by destiny separated.

Another anecdote told by Thiébault (V: 31) deserves to be repeated briefly though it is more characteristic of Lambert than of Lagrange. Thiébault having asked the former who were the leading mathematicians, Lambert answered somewhat like this "Euler and d'Alembert are

---

[33] In 1735 Euler suffered a severe illness and lost the use of one of his eyes; this misfortune was ascribed to overwork and to the climate of St. Petersburg. However, thousands of men have worked themselves to death and millions have lived in that city without losing their eyesight. The cause was deeper. After his return to St. Petersburg in 1766 a cataract formed in his good eye; by 1771 he was completely blind (Pasquier, 1927: 92–94).

[34] *Nouvelle Biographie Générale* XLV: 154–156, Paris, 1866; A. de Backer and C. Sommervogel, *Bibliothèque de la Compagnie de Jésus* VII: 1969, Paris, 1896. The général baron Thiébault, who wrote memoirs dealing with the years 1795–1820, a good source for the history of the French Revolution and the Empire, was Dieudonné's son. There are no references to Lagrange in the general's memoirs.

first ex aequo, Lagrange is third, I am fourth, and there is no need of going any further as I cannot think of anybody else worth quoting after these."

During Lagrange's stay in Berlin he had devoted himself primarily to mathematical and mechanical investigations published by the Academies of Berlin, Turin, and Paris. However, he found time to prepare what may be called his first treatise to appear in book form. Euler had published a treatise on algebra in 1770,[35] a French translation of it was made by Jean III Bernoulli (1744–1807) and published under the title *Elémens d'algèbre* (2 v., Lyon, 1774); the second volume of that French translation contains a very long supplement on indeterminate analysis by Lagrange (p. 369–684).[36]

Inasmuch as Lagrange wrote long letters to d'Alembert from the time of his arrival in Berlin until his friend's death (1783), that is during seventeen years, we might expect to find in those letters much information on his Berlin life. Such expectations are disappointed. The letters written in Berlin might have been written just as well from Shangri-La. There is no appearance that Lagrange ever travelled in Germany, except to go from Paris to Berlin and back; there is no appearance even that he explored the city. Apart from a little (very little) academic gossip, there is no Berlin news. It was the same everywhere. When he spent a few days in London as the Marchese Caràccioli's guest, on his way from Paris to Berlin, he wrote d'Alembert (London, September 23, 1766; XIII: 78). "The marquis Caràccioli at whose house I am staying does everything to make my sojourn here as pleasant as possible; he is taking and showing me everywhere, but I have not enough curiosity to appreciate correctly his attentions." That sentence is as revealing as can be. We might say that Lagrange never lived in Germany, that he never lived in Berlin, though he spent there twenty-one years, the richest years of his life.

Thanks to his reputation, to d'Alembert's friendship and to his own gentleness he remained *persona grata* to the king, but it is not probable that he saw much of the latter. The king must have perceived quickly that there was not much to be got out of his best mathematician except memoirs which everybody said were admirable, but which nobody could read. Lagrange liked his position and was anxious to keep it. Ten years after his arrival he was afraid that if Frederick should die, he himself might fail to obtain the favor of the new king, and he wrote to d'Alembert (September 25, 1776; XIII: 323) asking him to prepare the ground. This shows how prudent he was and farseeing. As a matter of fact, Frederick the Great lived ten years longer but Lagrange had been right in suspecting that Berlin would cease to be a comfortable residence for him after the loss of his royal patron.

When the death of Frederick the Great appeared imminent the chancelleries of Europe began to speculate as to what would happen when he ceased to exist. Charles Alexandre de Calonne (1734–1802), Minister of Finances under Louis XVI, dispatched an agent to Berlin to investigate. That agent was no less a person than the Count of Mirabeau (1749–1791), who was to play so momentous a role during the first years of the French Revolution. Frederick died on August 17, 1786, and Mirabeau witnessed the difficulties which were bound to follow almost immediately. The new king, Frederick William II (1786–1797), was entirely unworthy of the uncle, Frederick the Great, whom he had the perilous honor of succeeding. The critical period through which Prussia was then passing was aggravated by his pusillanimity. His administration was reactionary and futile; partly as a reaction against the rationalism and cosmopolitanism of the great Frederick it favored anti-foreign (especially anti-French), and mystical, theosophical tendencies. Such an atmosphere must have been very distasteful to Lagrange. Mirabeau was aware of the latter's troubles and wrote the French government, suggesting that the great mathematician be invited to establish himself in Paris. The following passage taken from his letter of November 28, 1786, deserves to be quoted, because it throws more light on Lagrange's personality.[37]

> If I am not mistaken, there is here, at this moment, an acquisition to be made, worthy of the King of France, and M. de Calonne is the very man who ought to lay the proposal before His Majesty. The illustrious La Grange, the greatest mathematician that has appeared since Newton, and who, by his

[35] *Vollständige Anleitung zur Algebra*, 2 v., St. Petersburg, 1770.

[36] I have seen only the second edition (2 v., Lyon, 1795). Both the German text and Lagrange's supplement are reprinted in Euleri *Opera Omnia* ser. 1, I, Leipzig, 1911. Lagrange's supplement was also reprinted in Lagrange's *Oeuvres* VII: 5–180, 1877.

[37] Mirabeau, *Secret history* (II: 96–100, 1895); for the French original (II: 97–99, 1789).

understanding and genius, is the man in all Europe who has most astonished me; La Grange, the most sage, and perhaps the only true practical philosopher that has ever existed; worthy to be commended for the pertinacious calmness of his mind, his manners, and his conduct; in a word, a man affectionately respected by the small number of men whom he would admit to be of his acquaintance; this La Grange has lived twenty years at Berlin, whither he was invited, in his youth, by the late King, to succeed Euler, who had himself pointed him out as the only man proper to be his successor. He is much disgusted, silently but irremediably disgusted, because that his disgust originates in contempt. The passions, brutalities, and lunatic boastings of Hertzberg;[38] the addition of so many with whom La Grange cannot, as Academicians, without blushing, associate; the very prudent dread of seeing himself held in painful suspence, between the philosophic repose which he regards as the first good, and that respect which he owes himself, and which he will not suffer to be insulted; all induce him to retire from a country where the crime of being a foreigner is not to be forgiven, and where he will not support an existence which will only be tolerated. It cannot be doubted but that he would willingly exchange the sun and the coin of Prussia for the sun and the coin of France, the only country on earth where men pay homage to the genius of science, and confer lasting fame; the only country where La Grange, the grandson of a Frenchman, and who gratefully recollects that we have made him known to Europe, would delight to live, if he must renounce his old friends and the abode of his youth. Prince Cardito di Laffredo, ambassador from Naples to Copenhagen, has made him the handsomest offers, in the name of his Sovereign. He has received pressing invitations from the Grand Duke and the King of Sardinia. But all these proposals would easily be forgotten, if put in competition with ours. And will not the King of France likewise, aided by a worthy comptroller-general, at the time when he would extend that empire of benevolence which appertains to him alone—would not the King of France endeavour to acquire a man whose merit is known to all Europe? La Grange here receives a pension of six thousand livres. And cannot the King of France dedicate that sum to the first mathematician of the age? Is it beneath Louis XVI, to invite a great man, from a miserable academy, who is there misunderstood, misallied, and thus, by the most noble warfare, to extirpate the only literary corps that has wrestled against his proper academies? Would not this act of generosity be superior to those that are usually performed? France, with pernicious policy, has been the asylum of princes, with whose necessities she was burdened. Why will she not welcome a great man, who would but add to her worth? Has she so long enriched others with her losses, and will she not enrich herself by others' errors? In fine, to speak of the minister I love, one De Boynes[39] has given eighteen thousand livres a year, for a useless place, to one Boscovich;[40] a man despised by all the learned of Europe, as a literary quack of poor abilities; and why will not M. de Calonne grant a pension of two thousand crowns to the first man in Europe of his class, and probably to the last great genius the mathematical sciences shall possess; the passion for which diminishes, because of the excessive difficulties that are to be surmounted, and the infinitely few means of acquiring fame by discovery?

I have the hope exceedingly at heart, because I think it a noble one, and because I tenderly love the man. I entreat I may have an immediate answer; for I own I have induced M. de la Grange to suspend his declarations on the propositions that have been made him, till he has heard what ours may be.[41]

Mirabeau's wise suggestion was acted upon and when the flattering invitation of the Paris court and Academy reached Lagrange, the latter was ready enough to accept it. Lagrange left Berlin on May 18, 1787 (XIV: 316).

X. MÉCANIQUE ANALYTIQUE

It is natural enough to interrupt for a moment our narrative of Lagrange's life and our analysis of his character in order to consider his greatest work, the *Mécanique analytique*, and the more so because that work largely composed in Berlin was completed, proofread, and published in Paris. It marks the transition between two parts of Lagrange's life, the Berlin and the Paris periods.

When we speak of the *Mécanique analytique* as his greatest work we should bear in mind that such a statement is not quite permissible. There is no common measure between his various achievements, and we cannot say that this or that one is absolutely more important than all the others. Yet the *Mécanique analytique* has

[38] Count Ewald Friedrich von Hertzberg (1725–1795), Prussian statesman and minister (G. S.).

[39] I assume this is the comte Bourgeois de Boynes, who was minister of the French navy and was replaced in 1774 by Turgot (G. S.).

[40] Dalmatian mathematician and Jesuit, Ruggiero Giuseppe Boscovich (1711–1787) (G. S.).

[41] Mirabeau added between brackets "on sent bien que lui qui est dans l'antre ne peut en faire [= ne peut faire de propositions] d'aucune espèce." This is wrongly translated "I need not repeat that he whose hands are tied must call for help." The correct meaning is "one feels that he who is in the den cannot make propositions." It is clear to me that Mirabeau and Lagrange had been conferring; Lagrange had explained that while he could not make offers without jeopardizing himself, he could and might accept good offers made to him.

MÉCHANIQUE ANALITIQUE;

Par M. DE LA GRANGE, de l'Académie des Sciences de Paris, de celles de Berlin, de Pétersbourg, de Turin, &c.

A PARIS,

Chez LA VEUVE DESAINT, Libraire; rue du Foin S. Jacques.

M. DCC. LXXXVIII.

AVEC APPROBATION ET PRIVILEGE DU ROI.

FIG. 5. First edition, 1788, quarto.

MEMORIA SUL PRINCIPIO DELLE VELOCITÀ VIRTUALI

DEL CAVALIERE

VITTORIO FOSSOMBRONI

ARETINO

UNO DEI QUARANTA DELLA SOCIETÀ ITALIANA, SOCIO DELL'ISTITUTO DI BOLOGNA ec.

FIRENZE MDCCXCVI.

PER GAETANO CAMBIAGI STAMPATORE GRANDUCALE.

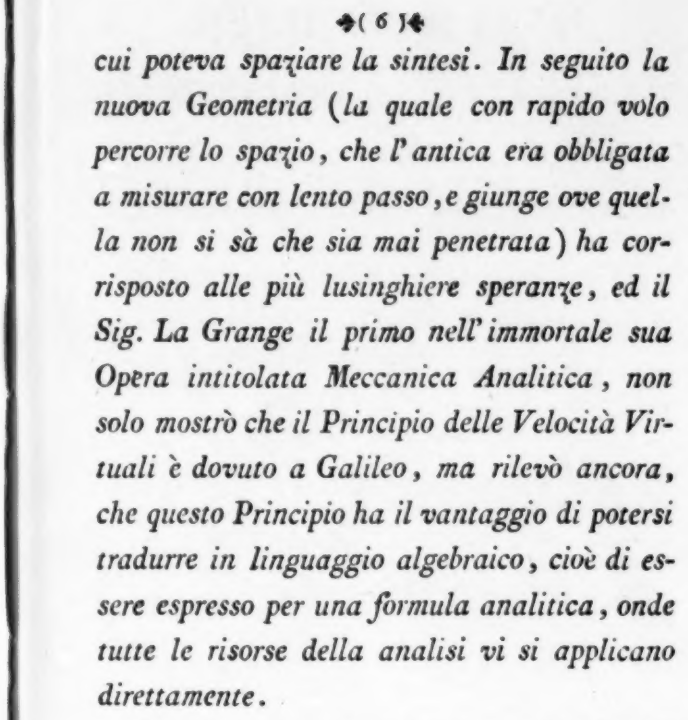
FIG. 6. Fossombroni's memoir on the principle of virtua velocities, Florence, 1796.

captured the imagination of most people, whether competent to appreciate it or not. The non-mathematicians were intrigued by its abstraction and by the complete absence of figures. The mathematicians realized that that book was much more than a new contribution to mechanics, it introduced a new method of almost unbelievable simplicity. It established a definite advance in the evolution of mechanics. It is one of the outstanding landmarks in the history of both mathematics and mechanics. Starting from the principle of virtual velocities, first announced by him in 1764 (VI: 8), Lagrange gave a simple means of writing immediately the fundamental equations of a problem. To repeat Mach's apt saying, "The mechanics of Lagrange is a stupendous contribution to the economy of thought."[42]

[42] Ernst Mach, *The science of mechanics*, 3rd. ed., 467, Chicago, Open Court, 1907.

The writing was done in Berlin and the manuscript was submitted to the Académie des Sciences not very long after the author's installation in Paris. The manuscript was examined by a commission the members of which were Laplace, Cousin, Legendre, and Condorcet; they concluded that it deserved to be published under the Academy's auspices. The report signed by Condorcet was dated February 27, 1788. The publication was made possible, because it was underwritten by the abbé Marie[43] and by Legendre. It was carried out very quickly, for Lagrange wrote to father di Caluso on April 28, 1788, that it had

[43] The abbé Joseph François Marie (Rhodez, 1738–Memel, 1801) had been active in obtaining Lagrange's appointment in Paris (I: xxxvi). He was an astronomer as well as a mathematician, and had published in 1774 a *Traité de méchanique* (*Nouvelle Biog. Gén.* XXXIII: 737, 1860). Nielsen (1929: 177).



❧(6)❧

cui poteva spaziare la sintesi. In seguito la nuova Geometria (la quale con rapido volo percorre lo spazio, che l'antica era obbligata a misurare con lento passo, e giunge ove quella non si sà che sia mai penetrata) ha corrisposto alle più lusinghiere speranze, ed il Sig. La Grange il primo nell'immortale sua Opera intitolata Meccanica Analitica, non solo mostrò che il Principio delle Velocità Virtuali è dovuto a Galileo, ma rilevò ancora, che questo Principio ha il vantaggio di potersi tradurre in linguaggio algebraico, cioè di essere espresso per una formula analitica, onde tutte le risorse della analisi vi si applicano direttamente.

Quel Principio dopo inventato da Galileo era rimasto quasi negletto, come penderebbe inutile una grande spada, fino a tanto che non nascesse un braccio atto a brandirla. In fatti

FIG. 7. Page from that memoir praising Lagrange's achievement. Beautiful typography.

Analytiſche Mechanik

von

Herrn La Grange

Mitgliede des National-Inſtituts der Künſte und Wiſſenſchaften zu Paris, der Akademien der Wiſſenſchaften zu Berlin, Petersburg, Turin u. ſ. w.

aus dem Franzöſiſchen

mit einigen Anmerkungen und erläuternden Zuſätzen

von

Friedrich Wilhelm Auguſt Murhard.

Göttingen,

bey Vandenhoeck und Ruprecht 1797.

FIG. 8. German translation of the *Mécanique analytique* Göttingen, 1797.

just been completed (that particular letter has been quoted above.)[44]

After the publication of his great work, Lagrange was very depressed and for a long time unable to continue his mathematical investigations or at any rate disinclined to attempt any. It is said that for two years he did not open his copy of the *Mécanique analytique*, but that is probably an exaggeration. Lagrange was mentally exhausted; it is possible that he was less inclined to work in Paris than he had been in his Berlin isolation. It is tempting to compare Lagrange's exhaustion with that suffered by Newton; the latter's illness, however, did not become alarming until 1693, that is six years after the publication of the *Principia*, while Lagrange's apathy followed immediately the publication of the *Mécanique*. It is interesting by the way to realize that those two immense achievements occurred almost exactly with a century's interval (1687–1788).

The *Mécanique* was not translated into Italian, there was no need of that, but Vittorio Fossombroni published in 1796 a memoir on the principle of virtual speeds in which he spoke very highly of Lagrange "immortal work." This small quarto (192 pp., 7 fig.) beautifully printed (figs. 6, 7) is a proof of the eagerness with which Italian mathematicians followed the career of their illustrious countryman; it must have helped to spread the new mechanics; in spite of the personality of its author it has been strangely overlooked by historians (such as Cantor, Riccardi, Loria). The Aretinian Fossombroni, born in 1754, was a mathematician and economist; when his book on Lagrangian mechanics appeared he was minister of foreign affairs in Tuscany. Later he was like Lagrange a senator and count of the

[44] According to Murhard's German translation (ix, Göttingen, 1797) the *Méchanique analitique* was published on May 4. However, Lagrange may have distributed what would be called today pre-publication copies.

MÉCANIQUE ANALYTIQUE,

Par J. L. LAGRANGE, de l'Institut des Sciences, Lettres et Arts, du Bureau des Longitudes; Membre du Sénat Conservateur, Grand-Officier de la Légion d'Honneur, et Comte de l'Empire.

NOUVELLE ÉDITION,

REVUE ET AUGMENTÉE PAR L'AUTEUR.

TOME PREMIER.

PARIS,

M^me^ V^e^ COURCIER, IMPRIMEUR-LIBRAIRE POUR LES MATHÉMATIQUES.

1811.

FIG. 9. Second edition, Paris, 1811.

THÉORIE DES FONCTIONS ANALYTIQUES,

CONTENANT

LES PRINCIPES DU CALCUL DIFFÉRENTIEL,

DÉGAGÉS DE TOUTE CONSIDÉRATION

D'INFINIMENT PETITS OU D'ÉVANOUISSANS,

DE LIMITES OU DE FLUXIONS,

ET RÉDUITS

A L'ANALYSE ALGÉBRIQUE

DES QUANTITÉS FINIES;

Par J. L. LAGRANGE, de l'Institut national.

A PARIS,

DE L'IMPRIMERIE DE LA RÉPUBLIQUE.

Prairial an V. 1797

FIG. 10. First edition, Paris, 1797.

empire, but unlike Lagrange he outlived that empire, and after Napoleon's downfall became Prime Minister of Tuscany. He was best known for his saying "Il mondo va da sè." He died in 1844 (*Enc. ital.* XV: 776, 1932).[45]

A German translation was made and annotated by Friedrich Wilhelm August Murhard and published in Göttingen in 1797. This Murhard was extraordinarly young when he made it, for he was born in Cassel in 1779 or 1778. He studied mathematics in Göttingen under Abraham Gottlieb Kaestner and this explains his deep interest in the history of science, an interest which was perhaps more common at the end of the eighteenth century than it is now. He wrote the first history of the calculus of variations (1796), a bibliography of magnetism (1797), a history of modern physics (1798–99) and published five parts of a physico-mathematical bibliography (1797–1805).[46] He travelled in the Near East and published books on Constantinople (1804) and the Greek archipelago (1807–1808). Later his attention was turned more and more to national and international politics and he became a student of government and a publicist. He died in his native city in 1853. See Cantor's note in *Allg. deutsche Biographie* XXIII: 62, 1886, and Cantor's *Vorlesungen*.

Toward the end of his life Lagrange prepared a new, enlarged, edition of his *Mécanique analytique*

[45] There is a single letter of his to Lagrange dated Florence September 14, 1798 (XIV: 294).

[46] F. W. A. Murhard, *Specimen historiae atque principiorum calculi quem vocant variationum . . .*, Göttingen, 1796; *Versuch einer historischchronologischen Bibliographie des Magnetismus*, Kassel, 1797; *Geschichte der Physik*, I, in 2 pts., Göttingen, 1798–1799; *Bibliotheca Mathematica oder Literatur der mathematischen Wissenschaften*, 2 pts., Leipzig, 1797–1798; *B. M. oder Literatur der mechanischen und optischen Wissenschaften* pts. 3–5, Leipzig, 1803–1805.

THEORICA
DAS FUNÇÒES ANALYTICAS,
QUE CONTEM
OS PRINCIPIOS DO CALCULO DIFFERENCIAL.
LIVRES DE TODA A CONSIDERAÇAÕ DE QUANTIDADES
INFINITAMENTE PEQUENAS OU DE DESVANECENTES,
DE LIMITES OU DE FLUXÕES, E REDUZIDOS
Á ANALYSE ALGEBRICA DAS QUANTI-
DADES FINITAS.
POR M. LA GRANGE
E DE ORDEM DE
SUA ALTEZA REAL
O PRINCIPE
NOSSO SENHOR
TRADUZIDA DO FRANCEZ
POR
MANOEL JACINTO NOGUEIRA DA GAMA
Cavalleiro Profeſſo na Ordem de S. Bento de Aviz, Bacharel Formado em as Faculdades de Mathematica, e Philoſophia pela Univerſidade de Coimbra, Capitaõ Tenente da Armada Real, e Profeſſor de Mathematica na Academia Real da Marinha.
PRIMEIRA PARTE.

LISBOA,
Na Offic. de JOAÕ PROCOPIO CORREA DA SILVA,
Impreſſor da Santa Igreja Patriarcal.
ANNO M. DCC. XCVIII.
Impreſſo por ordem da Sua Mageſtade.

FIG. 11. Portuguese translation, Lisbon, 1798.

Lagrange's
Theorie
der
analytiſchen Funktionen,
in welcher die Grundſätze der Differentialrechnung vorgetragen werden, unabhängig von Betrachtung der unendlich kleinen oder verſchwindenden Größen der Grenzen oder Fluxionen, und zurückgeführt auf die algebraiſche Analyſis.

Aus dem Franzöſiſchen überſetzt
von
Johann Philipp Grüſon
Königl. Profeſſor der mathematiſchen Wiſſenſchaften am adeliſchen Cadettencorps in Berlin, und ordentlichem Mitgliede der Königl. Preuß. Akademie der Wiſſenſchaften.
Schöner.

Berlin,
bey F. T. Lagarde
1798.

FIG. 12. German translation, Berlin, 1798.

and his last years were devoted to that task. The treatise was so much expanded that it had to be printed in two volumes, the first of which, containing the statics and the beginnings of dynamics (sections I–VI), appeared in 1811.[47] The first sheets of volume 2 were already printed when Lagrange died (1813); the edition was continued by his colleague, Prony, and by Garnier, professor at the Royal military school. Section IX (movement of rotation) had only been sketched out by Lagrange; it was elaborated by Prony, J. Binet, and Lacroix. The second volume completing the dynamics (section VII to XII) finally appeared in 1815.

From what we have just said it is thus clear that Lagrange had not finished the manuscript of that volume before his death, for section IX was only begun and Bertrand found later (1853) that section VIII contained many errors of calculation which he was obliged to remove when preparing the third edition. Virey states (1813: 17) that Lagrange was contemplating the writing of a third volume devoted to celestial mechanics. That is not impossible but considering the incompleteness of volume 2 and Lagrange's waning forces, it is not probable that he made any effort

[47] Lagrange presented that volume to the Académie des Sciences on September 16, 1811 (*Inst. de France, Procès-verbaux* 4: 534, 1913.)

toward the preparation of that volume. His profound studies in celestial mechanics published in the form of monographs by the academies of Berlin and Paris helped his younger colleague, Laplace, to built up a new monument, more impressive than the *Mécanique analytique* though less original, the *Traité de Mécanique céleste* (5 v. Paris, 1798–1825).[48]

It is interesting to note that when Lagrange published the first volume of the second edition in 1811 he felt obliged to introduce the following remark in his preface (XI: xiv).

> On a conservé la notation ordinaire du Calcul différentiel, parce qu'elle répond au système des infiniment petits, adopté dans ce Traité. Lorsqu'on a bien conçu l'esprit de ce système, et qu'on s'est convaincu de l'exactitude de ses résultats par la méthode géométrique des premières et dernières raisons, ou par la méthode analytique des fonctions dérivées, on peut employer les infiniment petits comme un instrument sûr et commode pour abréger et simplifier les démonstrations. C'est ainsi qu'on abrège les démonstrations des Anciens par la méthode des indivisibles.

In other words, after more than a century of efforts by the leading mathematicians of Europe, the calculus was not yet sufficiently well established in 1811 to inspire general and implicit confidence. The task of making the calculus rigorous was begun anew after Lagrange's death by Augustin Cauchy (1789–1857).

To return to the editions of the *Mécanique analytique*, a third one was prepared by Joseph Bertrand (2 v. Paris, 1853–1855), including mathematical corrections by Bertrand, mathematical appendices by Poinsot, Lejeune-Dirichlet, Bertrand, V. Puiseux, J. A. Serret, Ossian Bonnet, A. Bravais; three other memoirs by Lagrange and four notes taken from his manuscripts (II: 367–381); a list of Lagrange's published writings and manuscripts compiled by S. F. Lacroix.[49]

A fourth edition prepared by Gaston Darboux and forming volumes XI–XII of the *Oeuvres* appeared in time (1888–1889) to celebrate the hundredth anniversary of the first. It reproduces the third one with all its mathematical appendices, plus two new ones by Darboux, but minus the Lagrange memoirs already reprinted in other volumes of the *Oeuvres*. It is typical that with the exception of the bibliographical notes of Lacroix neither of those editions includes historical information, such as a mathematical historian would have added in many places. The editors have not even taken the elementary trouble of indicating the parts of the text belonging to the original edition of 1788 and representing Lagrange's first thoughts on the subject. It is clear that they had no idea whatsoever of the responsibilities which the edition of a classic implies.

Note on the title of the first edition.—The title of the first edition "Méchanique analitique" (repeated by Condorcet in the academic placet) is very puzzling, for the two words are misspelled, and they are misspelled in opposite ways, the first being an archaism and the second a modernism! The *h* in the first word is a vestige of the Greek *chi*, while the first *i* in the second word is a rejection of the Greek *upsilon*. The spelling méchanique was frequent in the eighteenth century, but as it conflicted with the pronunciation the *h* tended to disappear.[50] In the *Abrégé du Dictionnaire de Trévoux* (Paris, 1762) we still have méchanique, and the textbook of the abbé Joseph François Marie was entitled *Traité de méchanique* (Paris, 1774), but in the *Dictionnaire de l'Académie Françoise* (ed. of Lyon, 1777), we find mécanique. The evanescence of that Greek *h* is well illustrated in the *Traité de dynamique* of d'Alembert, where we find "méchanique" in the dedication and in the beginning but "mécanique" further on (e.g., p. 50, 68)!

The word analitique is more puzzling still. Why did Lagrange abandon the "*y* grec" of the word analytique and keep it throughout the book in other words such as métaphysique, système, hyperbolique, hydrodynamique? In this respect the early dictionaries gave him no countenance. In every one consulted by me I found analyse, analytique, etc.

Did Lagrange ignore the Greek origin of the word analyse? or was he influenced by the Italian word analisi? Examples of Italianisms in his writings are very rare,[51] by the way, and this proves that his use of the Italian language was exceptional.

[48] Volumes I–IV appeared within Lagrange's life time, 1798–1805, volume V only in 1823–1825. We do not know Lagrange's reactions to these volumes because his correspondence with Laplace begun in Berlin 1773 stopped after his arrival in Paris.

[49] Lacroix's list of Lagrange's published writings was already included in the second edition (II: 372–378, 1815).

[50] On the contrary, the *h* remained in "machine," because it agreed with the pronunciation.

[51] I can quote only two; "notices" in the sense of news (notizie) in a letter to d'Alembert (XIII: 242), and prémier for premier in the dedication of the *Miscellanea taurinensia* (see our facsimile).

Some readers will say "What does it matter?" Well, is it not strange that one of the outstanding monuments of the human spirit should be labelled in such an erratic way? It does not matter very much perhaps, yet it does matter.

XI. HIS TWO MARRIAGES

Lagrange was twice married, and his two marriages, being similar in essential respects, help us to understand his personality.

His first marriage took place in Berlin within a year after his arrival in that city. He married his cousin, Vittoria Conti, in 1767.[52] We know the circumstances of that marriage pretty well from his correspondence with d'Alembert.

The latter wrote on September 21, 1767 (XIII: 100)

> My dear and illustrious friend, they write from Berlin that you have taken the plunge [53] and married one of your cousins whom you have sent for in Italy; accept my compliments, for I think that a great mathematician must be able before all to calculate his happiness and after having made that calculation you found marriage for solution.

Lagrange answered on November 20, 1767 (XIII: 102)

> My dear and illustrious friend, I have received your letters and compliments and thank you wholeheartedly. I do not know whether I have calculated well or badly, or rather I think that I did not calculate at all, or else I would perhaps have done like Leibnitz, who reflected so much that he was never able to reach a decision. Whatever the case be, I will confess to you that I have never had any taste for marrying and that I would not have pledged myself if circumstances had not somehow obliged me to do so. Being in a foreign country, without friends and without connections, and with a delicate health, I have thought of inviting one of my relatives, whom I had known for a long time and with whom I had already lived for a few years in my father's house, to come and share my fate and take care of me and of all the things concerning me. That is the exact history of my marriage. I had not informed you of it, simply because the matter seemed to me of so little importance in itself, that it was not worthwhile to speak of it to you.

That was clear enough and d'Alembert replied in the same vein, on January 18, 1768 (XIII: 104):

> Inasmuch as your marriage, my dear and illustrious friend, is a matter of arrangement and convenience, I send you my compliment even as the priests give the absolution quantum possum et quantum tu indiges, and I find that, everything calculated and as a last result, you have done very well.

Lalande,[54] who was jealous and vain, wrote "Lagrange likes to thesaurize. He called to Berlin a cousin of his for whom he bought ribbons himself in order to reduce her expenses. He was obliged to marry her" (XIII: 100). We may overlook Lalande's ugly thoughts. Lagrange's explanation given above, unflattering to himself as it is, may be accepted as genuine. It is confirmed in a letter which he wrote to d'Alembert two years later, on July 15, 1769 (XIII: 142):

> My present situation is such that there is nothing more to be desired; it is true I am married, but I have no children and I have no wish of having any; my wife who is one of my cousins and has even lived for a long time in my family, is a very good housekeeper and without any pretension, so much so that everything considered my marriage is not a burden to me.

His candor seems almost cynical to us, but it was not; he did not feel that there was anything wrong in his behavior. According to Virey (p. 8) his wife was the victim of a slow and cruel illness. There are no references to this in Lagrange's letters. The one in which he probably announced her death is unfortunately lost, and we have only d'Alembert's answer to it dated Paris, September 27, 1783 (XIII: 377). D'Alembert himself was very ill and this was his last letter [to Lagrange].

> My dear and illustrious friend. I am so weak that I have not the strength to write and can just dictate a few words. I take the tenderest part in your misfortune, and what you say of it has touched me to the bottom of my soul. . . .

D'Alembert died a month later, on October 29, 1783.

In 1787, Lagrange moved to Paris, but he did not marry again until 1792, that is almost ten years after the death of his first wife. His second

---

[52] Her name is given in the *Primo secolo della R. Acc. di Torino*, 90, 1883, and by Loria (1937: 306). Investigations to find the marriage certificate in the archives of Berlin and Turin have failed.

[53] . . . vous avez fait ce qu'entre nous autres philosophes nous appelons le *saut périlleux*.

[54] Jérôme Lalande (1732–1807). French astronomer. See notice by Delambre (*Biog. univ.* XXIII: 215–232, Paris, 1819). It was this Lalande, who edited volumes 3 and 4 of Montucla's *Histoire des mathématiques* in 1802 (*Osiris* **1**: 550).

marriage has been very misleadingly described by mercenary writers who prefer to be entertaining rather than to be truthful. It is so much easier to crack a joke than to find the truth, and the reader may prefer the joke. The joke being that the old man married a very young woman who obliged him to go to balls, etc.; you recognize one of the commonplaces of comedy, a good opportunity for the author who must write so many pages a day to feel that he has really done something.

The facts are these: On May 31, 1792, he married (Renée Françoise) Adelaïde, daughter of his colleague the astronomer, Pierre Charles Le Monnier. Lagrange was then fifty-six which is not yet old age, and Adelaïde could not be so very young, for her father was twenty-one years older than Lagrange. Le Monnier had three daughters, Adelaïde being the second; the youngest daughter married a man nineteen years older than Lagrange, to wit, her own uncle, Louis Guillaume Le Monnier.[55] I do not know how young[56] and pretty Lagrange's new bride was, but I am sure she was sensible and domestic; otherwise he would not have married her. Lagrange's only justification of marriage we have seen, was as a means of simplifying his life. Therefore, he was careful each time to marry a woman who could be trusted not to cause difficulties and complications, a woman of his own kind. In the first case, he married a cousin whom he had intimately known for years; in the second case, a woman belonging to the academic family. Her grandfather, father, and uncle were members of the Paris Academy! At the time of his marriage he told Delambre (I: xlvii) "I have had no children from my first marriage, I do not know whether I shall have any from the second, I do not wish for any." This suggests that Adelaïde was still young enough to bear children; it is probable that, as his strength was then at a low ebb, he looked older than he was; it is equally probable that his bride looked younger than her years. What he needed most was a good companion, and there is good reason to believe that Adelaïde was that to him. He went out oftener with her to academic evenings and concerts, but seldom to the theatre, for he disliked the vitiated atmosphere of such places, and preferred to be at home or in the open air. His second marriage lasted five years longer than the first, twenty-one years against sixteen.

Both marriages were what the French call marriages of convenience, and this was aggravated by the fact that Lagrange wanted no children. He may seem very selfish, and maybe he was, but perhaps he had no choice for he was the slave of his destiny. He offered his wives no romance, no passionate love, but a share of his fame and a genuine affection deepening with the passing years. Most women want much more and get much less. There are perhaps a good many others who would want nothing better than the simple loyalty of a good and great man, the security of his home, and the tender affection of his heart.

Lagrange had no children of his own but the family was continued by his younger brother, practising law in Turin. The latter's daughter married the Baron Giovanni Plana (1781–1864) of Voghera, Lombardy, who was Professor of Mathematics at the University and Director of the Observatory of Turin. Plana's main work was published in French, *Théorie du mouvement de la lune* (3 v., Turin, 1832).[57] In 1868, there were living in Torino a Sig. Cavaliere Michele Lagrange, retired colonel, and the Signora Baronessa Sofia Plana,[58] who was probably Plana's widow, that is, Lagrange's own niece. In a letter dated Genoa, July 14, 1941, Professor Gino Loria wrote me that descendants of the Lagrange family were still living in Turin, and he offered to put me in touch with them; soon afterwards it became impossible to communicate with him because of the war.

XII. HIS FRIENDSHIPS

Lagrange's first mathematical correspondents were Fagnano and Euler, but his first friend was d'Alembert. That was the main friendship of

[55] Pierre Le Monnier (1675–1757), astronomer, had two sons, Pierre Charles Le Monnier (1715–1799), astronomer, and Louis Guillaume Le Monnier (1717–1799), first physician to Louis XVI, botanist. The Monniera (a Rutaceae) was named after the latter in 1758, having been discovered by Per Löfling in Guiana. The father and two sons were all members of the Academy of Sciences. Pierre Charles was Adelaïde's father, Louis Guillaume was her uncle and brother-in-law. Biographies of the three in *Biog. univ.* XXIV: 62–65, 1819.

[56] Vain efforts have been made at Loria's request to find the marriage certificate, and by me to find her obituary, which might have mentioned her age. I do not even know how long the Countess Lagrange survived her husband.

[57] He was a member of the Paris Academy and his biography was read by Elie de Beaumont in 1872. See also the *Primo secolo della R. Acc. di Torino*, 119–122, 1883. *Ecole Polytechnique, Livre du centenaire* III: 606–608, 1897.

[58] Boncompagni's *Bullettino* **1**: 183, 1868.

his life. He admired[59] and respected Euler but did not love him; as far as we know he made no friends in Germany. His friendship with d'Alembert was an intimate but distant one. They became acquainted in Paris in 1763 to 1764; Lagrange visited Paris again in 1766, then proceeded to Berlin and never came back until after d'Alembert's death. Hence, their physical experience of one another was very restricted. Yet d'Alembert had taken a liking to the younger man (Lagrange was nineteen years younger than himself), their letters always began and ended with very friendly words, and they were anxious about one another's health. Extracts from their letters quoted above have already given the reader some idea of their intimacy.

Euler was the first foreign mathematician to encourage Lagrange, and he treated the unknown youth with great consideration and generosity. They corresponded irregularly for twenty years but never met. Indeed, Euler left Berlin in May 1766, while Lagrange arrived there only in November. Irrespective of the difference of age (thirty years) they were too different to harmonize. Euler was deeply religious in an old fashioned way, he was a *pater familias*, almost a patriarch. Lagrange was none of that. Their only points of contact were mathematical and though that might suffice for emulation, it did not for mutual love. The *Lettres à une princesse d'Allemagne* (1768)—a reaction against the Voltairian ways of Frederick the Great and his circle—obtained an immense success,[60] but they naturally displeased the "philosophes." Lagrange wrote to d'Alembert on June 2, 1769 (XIII: 132), that it would have been better for Euler's honor not to publish them, and d'Alembert answered a fortnight later (XIII: 135) "From what you say those letters are his Commentary on Revelation.[61] Our friend Euler is a great analyst, but a poor philosopher." A few months later, August 7, 1769 (XIII: 147) he added

> You were quite right to say that it would have been better for his honor not to print that work. It is unbelievable that as great a genius as his in Geometry and Analysis proves himself inferior to a small schoolboy in Metaphysics, not to say, so flat and absurd, and we may well repeat the proverb, *Non omnia eidem Dei dedere.*

Their common reaction against Euler's philosophy reveals one of their bonds of sympathy. Another one was their devotion to mechanics. It was Lagrange's destiny to complete in his *Mécanique analytique* (1788) some of the ideas which d'Alembert had announced in his *Traité de dynamique* (1743). However, the main bond was that mysterious something, their liking one another from the beginning until Death took one of them away.

After his first visit to Paris and his first meeting with d'Alembert, he wrote from Turin, May 30, 1764 (XIII: 10):

> My dear and respectable friend, I passed through Geneva as I had planned to, and through the favor of your introduction, I have had the honor of dining with M. de Voltaire, who has received me very graciously. He was that day in high feather, and his jokes were always aimed as usual at religion, which much amused the whole company. He is truly a character who deserves to be seen (*C'est en vérité, un original qui mérite d'être vu*). . . .

The letter ends

> Adieu my dear Sir, be well and keep me your precious friendship, which I consider as the main advantage which I have derived from my journey to France. I love you as much as I esteem you, and those feelings will always be those of your very humble and very obedient servant.

Much of that is conventional, but there is something more which must have touched the older man. Lagrange often ends with the salutation "Adieu, fare well and love me as I love you," that is, of course, the old Latin formula, *Vale et me ama*—but he adds (1771; XIII: 224) "I kiss you a million times with my whole heart," which is less usual in mathematical correspondence, even in the eighteenth century.

The letters which they exchanged for twenty-four years, fill a whole volume (XIII) of the *Oeuvres*. The subjects dealt with are primarily and almost exclusively mathematical. This reveals a profound difference between the two friends. Lagrange had the reputation of being interested in many things, but as far as we can

---

[59] To people who asked for guidance Lagrange used to say "Etudiez Euler!" "Study Euler if you wish to become a geometer, and try to solve the queries which he is asking himself" (Maurice, 1819: 171).

[60] There were twelve French editions, plus nine in English, six in German, four in Russian, two in Dutch, two in Swedish, one each in Italian, Danish, and Spanish!

[61] Reference to Newton's Biblical studies probably anterior to 1693. His *Observations upon the Prophecies of Daniel and the Apocalypse of St. John* were published posthumously in London in 1733, a Latin translation appeared in Amsterdam in 1737, and a German one in Leipzig in 1765.

judge from his correspondence his interest in most things was too weak to find expression. In this respect the scope of Lagrange's thought was far more narrow not only than d'Alembert's, but even than Euler's. Lagrange is one of the best examples of the pure mathematician; this does not mean that he was a greater mathematician than the two older men, but they were greater men, or let us say, their minds were richer.

The contrast between Lagrange and d'Alembert is particularly striking. The latter was a man of letters and a humanist as well as a mathematician. It is possible to follow his literary thoughts in a long series of volumes; though his mathematical writings were his spring board to fame, he would still be famous if those writings did not exist. Lagrange did not lack literary imagination nor sensibility, but those delicate thoughts were overwhelmed and almost choked out of existence by the mathematical cogitations dominating his mind and brooking naught else. Lagrange's letter of March 20, 1779 (XIII: 347), wherein he thanks his friend for a copy of the latter's *Eloges* is significant. "It is not an outsider's business to appreciate such productions. I can only feel [their merit], and I assure you, I am the more penetrated with it that it is beyond my power. . . ." This letter is one of his best, it proves that his sensibility was not by any means dead. Some of his eulogists have ascribed to him interests in many things, interests which probably existed, but in such passive and latent forms that he deserves no praise for them. After all, we can give him credit only for the ideas which he expressed and the feelings which he substantiated.

To illustrate consider the reactions of those two illustrious men to music. Lagrange once explained his own very clearly to Delambre (I: xlviii) "I love music because it isolates me; I listen to the first three bars, with the fourth one I am already lost; I give myself up to my own reflections, nothing interrupts me and in that way I have solved many difficult problems." Strange love indeed! Think of a man who would say "Dear Mary, I love you so much, for I cannot see you without thinking of sweet Beatrice. . . ." On the contrary, d'Alembert's love of music was deep and genuine, he loved music for its own sake, and the quarrel which was then dividing the musical public of Paris meant a great deal to him. Although he was a Frenchman and very proud of it, he took up the cudgels for the Italian musicians against the insipid followers of Lully and Rameau.[62]

One could not write the history of arts and letters, say, the history of humanities, in the eighteenth century without referring repeatedly to d'Alembert; one would have to refer to Euler at least once because of his letters to a German princess, but there would be no need whatsoever of mentioning Lagrange.[63]

D'Alembert, who had received copies of his young friend's papers as they appeared, had caused them to be bound in four volumes. He gave that precious set to Condorcet on condition that the latter would eventually transmit it to a young geometer. Condorcet gave it to Lacroix, Lacroix to Biot. The latter enriched the collection which grew from four volumes to six and transmitted it to Jacques Binet. Though Binet was twelve years younger than Biot, he died before him and the books came back to Biot. In order to avoid a similar accident Biot gave them now to a man forty-two years younger than himself, Edmond Bour, but he almost failed again for Bour died only four years later than Biot, who lived to be eighty-eight. Bour bequeathed the volumes to the Academy which was asked to designate the new temporary owner. I do not know what happened next;[64] I suppose the books got stuck in the Academy's library.

It is more difficult to speak of Lagrange's friendships with other men. His most friendly messages were sent to Condorcet but their correspondence lasted only from 1771 to 1778, then to Laplace from 1773 to 1784, though the latter's vanity annoyed him. After his departure from Berlin he never wrote to his colleagues of twenty years! We have very few letters of the Paris period; making allowance for the loss of some of those letters, it is probable that he wrote fewer than in Berlin and did not begin extensive correspondences with anybody. He had no need of writing to his French colleagues whom he met very frequently at the Academy and elsewhere.

---

[62] See his essay "De la liberté de la musique" in his *Mélanges de littérature et d'histoire* IV: 383–462, Amsterdam, 1764. The more passionate quarrel, between the Piccinnists and the Gluckists, did not flare up until later. The *Iphigénie en Aulide* of Gluck dates from 1774, and Piccinni made his debut in Paris in 1776.

[63] There is no mention of Lagrange in the correspondence of Grimm, Diderot, Raynal, Meyster, etc., in Maurice Tourneux's edition (16 v., 1877–1882).

[64] Story told in the biography of Edmond Bour (1832–1866) in the *Livre du centenaire de l'Ecole Polytechnique* I: 147, 1895.

For example, there is good reason to suppose that he was friendly with Lavoisier, but no letters passed between them. Family letters or letters addressed to non-mathematicians are very few indeed, or at any rate, very few are known to exist.

Lagrange was too self-centered a man to make friends. His only true love was Urania.

XIII. VARIOUS TRAITS

One of the fundamental traits of Lagrange's character was his peacefulness. I would not call it kindness, for it was too passive. Lagrange did hardly ever hurt anybody and instinctively avoided every quarrel, or everything which might lead to a quarrel. In this again he differed materially from Euler and d'Alembert. Euler could be kinder or more actively kind, I believe, than Lagrange ever was, but on the other hand he was cantankerous and irascible. As to d'Alembert, he was eager to join in any controversy and have his say.

From the beginning Lagrange avoided controversies. In 1767, Alexis Fontaine des Bertins (1704–1771) attacked his views on the problems *de maximis et minimis*, but failed to draw his anger. Lagrange wrote to d'Alembert (August 2, 1769; XIII: 145)

> As I have a great distaste for quarrels and would be sorry to give pain to a man whose lights I respect and who has formerly honored me with his friendship, I have promised myself never to use the criticisms I have made of his work, unless I be forced to it for my own defense.

When Fontaine died very miserably, Lagrange wrote to the same (September 30, 1771; XIII: 212)

> I was very much affected by the death of M. Fontaine, and especially by the circumstances which accompanied it; though he flew out against me without rhyme or reason, my remembrance of his kind offices to me prevented me from being angry with him.

Daniel Bernoulli (1700–1782) criticized Lagrange by implication (XIII: 37), but the latter did not justify himself. The chevalier de Borda (1733–99) disagreed with him many times, but Lagrange answered very gently or kept quiet. D'Alembert having asked him what he thought of Borda's memoirs on fluids, he wrote (December 16, 1771; XIII: 223)

> I promise you to read them carefully and to give you my opinion, but only on condition that if my opinion is somewhat unfavorable, you will not get me into trouble with him, for I confess that I do not like quarrels and consider my peace of mind *rem prorsus substantialem.*

Borda had occasion to discover that Lagrange's gentleness was not a sign of weakness and after a discussion with him he told Delambre (I: xlix) "I am sorry to have to say this of a man like M. Lagrange, but I do not know any one who is more obstinate."

D'Alembert had quarrelled with Jérôme Lalande, who was somewhat of a schemer and difficult to deal with; he went so far as to write to Lagrange (June 18, 1765; XIII: 41) "That . . . of Lalande is gone to Italy; I hope the Italians will shake him like the vermin which he resembles," and in 1766 he called him (XIII: 76), "le . . . Lalande, barbet de Simpson."[65] Lagrange did not encourage those vituperations, but in 1771 he wrote to his old friend, "I am delighted that your reconciliation with Lalande is a true one; it seems to me that in every case peace is better than war."

In those days when journals were fewer in number than now, private letters had comparatively more importance, such letters were often circulated among friends and colleagues; everybody was aware of that and aggressive men used such vehicles to propagate their animosities. Lagrange was well aware of it too but restrained himself accordingly. His remarks concerning other people were always very moderate and cautious. D'Alembert was more outspoken and perhaps more sincere, Lagrange was more prudent.

This may be illustrated with reference to Lambert and Laplace, two mathematicians of considerable genius but vain and pushing. While Lagrange was in Berlin he was annoyed, as everybody was, by the former's self-conceit and churlishness, but he appreciated his merit and was anxious to help him improve his position (XIII: 142; Thiébault, 24–36). As to Laplace, Lagrange wrote to Concorcet (July 18, 1774; XIV: 27):

> I am a little surprized by what you say of M. de la Place; it is methinks the weakness of young men to be puffed up with their first successes, but presumption diminishes later in proportion as science increases.

[65] Simpson's spaniel. The English mathematician, Thomas Simpson (1710–1761) is probably meant.

The *Encyclopédie* had caused Diderot and his collaborators infinite trouble and although d'Alembert did not lack courage and did not fear controversies he had withdrawn from it about 1758. Lagrange would never have jeopardized himself in that company. When the historian, Carlo Denina (1731–1813) lost his chair in Turin because of an imprudent publication,[66] Lagrange expressed his sadness to an unidentified correspondent (July 11, 1778; XIV: 274) adding characteristically,

> I believe that in general one of the first principles of a wise man is to abide strictly by the laws of the country where he is living, even if some of those laws be unreasonable. Moreover, I have observed that the works which have brought upon their authors most contradictions and vexations, were not the kind of works which would be the most likely to give them a solid reputation, witness the *Encyclopédie* and many other French and Italian works. Our great Galileo owes his true glory only to his discoveries concerning motion, and the satellites of Jupiter. His famous Dialogues to which he owed all his misfortunes are the less good of his works and one can no longer bear to read them. Without them he would have enjoyed more happiness and might perhaps have become greater still by other discoveries.

The remark concerning Galileo is clearly wrong and condemns Lagrange, yet one can hardly blame the latter for defending his mathematical genius as well as he knew. It was not required of him to fight for justice and liberty, but it is well that other men like the "encyclopédistes" or Beccaria[67] were ready to do it.

His instinct for self-protection often took the form of a kind of blindness or callousness. It would not be fair to call that selfishness, for it is less a defense of the self than of the task to be done. Lagrange manages to get rid of everything which would hamper his own studies. When the Huguenot refugee Louis Dutens (1730–1812) planned an edition of Leibniz's works and discussed the project with his friend, Count Saluzzo, one of the originaters of the Turin society, the latter put him in touch with young Lagrange. Lagrange was induced to take charge of Leibniz's mathematical papers and to write an introduction to them. He gave some kind of a promise but did not keep it.[68] In Berlin, he seems to have remained very much aloof from his colleagues; in Paris he was more sociable, but the main part of his work was done, he was growing older and perhaps felt more keenly the need of relaxation.

The speech which he delivered at the Academy of Berlin on November 6, 1766, at the time of his reception, is remarkable for its laconism.[69] Compare it with one of d'Alembert's discourses, for example, the one he delivered when the King of Denmark visited the Académie des Sciences on December 3, 1768.[70] That comparison, however, is not completely to Lagrange's disadvantage. There was some grandeur as well as humility in that maiden speech.

Napoleon called him "la haute pyramide des sciences mathématiques"; perhaps it would have been better to compare him with an iceberg, which is as you know, so many times deeper than it is high. Moreover, the iceberg suggests coldness, and he was cold and distant. What a contrast again with d'Alembert's tenderness. Lagrange took refuge in silence and some people admired him the more for his sphinxlike attitude. Silence is golden, to be sure, but we cannot praise a man for the wise and brave words he did *not* say. We excuse his incapacity for small talk, but not so easily his silence at times when a few words from his lips might have encouraged or comforted braver men.

XIV. LAGRANGE'S RELIGION

One would be tempted to say, Lagrange had no religion; at any rate, I have not found in his writings the trace of religious feelings. If he had such feelings he did not express them. He had been baptized, but there is no reason to believe that either of his two marriages had been solemnized by the church, and he certainly died without having received the extreme unction.

There is no trace of anticlericalism in his writings. He simply was not interested in religious

[66] *Dell' impiego delle persone*, Florence, 1777.

[67] Cesare Beccaria of Milan (1738–1794) is meant of course, not Lagrange's teacher Giov. Batt. Beccaria of Turin (1716–1781). Those two Beccaria were not related. For the first see my article in *Bull. Hist. Med., Suppl.* no. 3: 283–308, 1944.

[68] *Leibnitii Opera omnia nunc primum collecta, in classes distributa, praefationibus et indicibus exornata studio Ludovici Dutens*, 6 v., Geneva, 1768. See Dutens' prefaces to volumes I and III. The preface to volume III begins "Equidem Praefationem Ludovicii De la Grange in hoc volumen pollicitus fueram: verùm quum aliis ille necessariis negotiis distentus promissis satisfacere nullo modo potuerit, has ego partes suscipere coactus sum, ac destituitur Literaria Respublica iis fructibus, quos ex praestantissimi hujus Mathematicii opera percepisset."

[69] Some two hundred words (XIV: 316).

[70] *Oeuvres de d'Alembert* III: 7–17, Paris, 1805.

questions. We must remember that he spent twenty years of his life under the patronage of Frederick the Great and that when he returned to Paris, the atmosphere of the circles which he frequented was definitely agnostic. After the Revolution, great efforts were made to reestablish the Catholic cult in France, but those efforts were resisted by many, in particular by members of the Academy of Sciences. A committee of that Academy composed of Lagrange, Laplace, and Monge waited in March 1801 upon the First Consul and told him that should he reestablish the cult he would lower his government and his age before Rome.[71] However, the First Consul was eager to pacify the country and a new concordat with Rome was signed on July 15, 1801 and solemnly published on Easter 1802.

When d'Alembert published his little book entitled *Destruction des Jésuites*[72] Lagrange wrote to congratulate him (Turin, July 6, 1765; XIII: 42), and he came back to the subject at least twice (XIII: 104, 115). In a letter to Condorcet (Berlin, February 29, 1774; XIV: 20) he remarked "I am delighted that you are finally rid of Boscovich; whatever the merit of his works I believe that they are always better than himself. He is a monk and Jesuit to burn" (*jésuite à brûler*). That is perhaps the least charitable sentence in all of his correspondence, and it shows that he was decidedly anti-Jesuit.

The Jesuits had been excluded from Portugal and her colonies in 1759. The Company of Jesus was suppressed in France in 1764,[73] and the Jesuits expelled in 1767. In the same year they were expelled from Spain and her colonies. Their general, Lorenzo Ricci (1703–1775, general from 1758) refused any kind of compromise declaring "*Sint ut sunt aut non sint.*" Finally, on July 21, 1773, the Holy Father,[74] being anxious to assure the peace of the church, issued the breve *Dominus ac Redemptor*, wherein the Company was suppressed, extinguished, abolished and abrogated everywhere and forever. It is one of the ironies of history that the Jesuit fathers found asylum in Prussia and Russia, thanks to the protection of the two free-thinking rulers Frederick II and Ekaterina II; their headquarters were even established for a time in Russia.

It was necessary to recall those facts in order to see Lagrange in correct perspective. Euler was deeply religious and theologically-minded (on the Protestant side), d'Alembert was anticlerical, Lagrange was anti-Jesuit as almost everybody was, but he was not anticlerical.

XV. LAGRANGE'S INTEREST IN SCIENCE AND THE HISTORY OF SCIENCE

Uncritical biographers have ascribed to Lagrange far more scientific curiosity than he really possessed. He was first and last a mathematician grudging the time and energy given to any problem but the mathematical ones which engrossed his mind. However, his fears concerning his own health had caused him early in life to obtain a superficial knowledge of physiology and medicine. That knowledge was gradually improved. It was he, and not Lavoisier, who was first to postulate that the real oxidation takes place not in the lungs but in the tissues throughout the body (Guareschi, 1913: 1–5). At the academical meetings which he attended with the greatest regularity he must have been thrown in contact with scientists of many kinds and must have listened to the discussion of every question of the day, yet we find very few references in his letters to scientific topics other than mathematical. Intelligent as he was he could not help obtaining a large amount of knowledge; his complete mastery in one field—the central field of science—made it easier for him to measure the progress made in the others and to realize the infinite complexity of the problems waiting for a solution. His knowledge and wisdom were sufficient to jeopardize whatever intellectual conceit might be left in him; he was modest not only as an individual but also as a member of the scientific profession. Toward the end of his life he admitted that our sciences are hardly more than glimmers in the dark abyss of our ignorance.

During the last quarter of the eighteenth century scientists paid considerable attention to the history of science. In general, they had no clear ideas of the intellectual requirements which a

---

[71] L. A. Thiers, *Histoire du Consulat et de l'Empire* III: 210, Paris, 1874. Of the three members of that committee, Monge was decidedly anticlerical, the two others were not. That commission is not mentioned in the *Procès-verbaux de l'Académie* 2, 1912.

[72] *Sur la destruction des Jésuites en France, par un auteur désintéressé* s.l., 1765. *Lettre à Mr * * * Conseiller au Parlement de * * * Pour servir de supplément . . .* s.l., 1767. *Seconde lettre*, s.l.a., July 15, 1767.

[73] What a blow this was to French education can be appreciated if one realizes that by that time the Jesuits controlled the best schools of France. For a list of Jesuit colleges in France in 1762 see *Centenaire de l'Ecole Normale* 3, 1895. Sarton, Montucla, *Osiris* 1: 519–520, 1936.

[74] G. V. A. Ganganelli, Franciscan, "Padre Lorenzo," Pope under the name of Clement XIV, 1769–1774.

good historian must meet, no conception of historical accuracy, but their historical curiosity cannot be denied, for a good many contemporary publications bear witness to it. Lagrange shared that uncritical curiosity. He read with great pleasure the history of the determination of the solar parallax written by Cassini de Thury (1772; XIII: 249). When Condorcet was planning to write a history of the sciences, Lagrange encouraged him (1773; XIV: 12). In July 1778 he expressed very foolish ideas on Galileo (XIV: 274). In January 1801 he criticized rightly the second part of Montucla's history then printing[75] and tried to obtain the history previously published (1796) in German by A. G. Kaestner (XIV: 296). The abbé Nicolas Halma's valuable edition and translation of the *Almagest* was finally published (2 v., 1813–1816) upon Delambre's and Lagrange's entreaties (XIV: 302).

What is more convincing, he made himself a number of historical investigations. Out of three memoirs on the determination of cometary orbits on the basis of three observations presented to the Academy of Berlin in 1778, 1783, the first was a historical introduction.[76] The *Mécanique analytique* (1788) contains interesting notes on the history of statics, hydrostatics, and dynamics, and the history of dynamics was enlarged in the second edition (1811). There are also brief historical introductions in his *Théorie des fonctions analytiques* (1797) and his *Résolution des équations numériques* (1798). It should be added that the value of those notes depends on Lagrange's personality rather than on their historical accuracy.

We may end this chapter with an anecdote told by François Arago in his biography of Denis Poisson (1781–1840).[77]

> "I am old," said Lagrange to Poisson, "during my long insomnias, I amuse myself by making numerical comparisons. This one may interest you; I want you to remember it.
>
> "Huygens was thirteen years older than Newton; I am thirteen years older than Laplace. D'Alembert was thirty-two years older than Laplace, and Laplace is thirty-two years older than yourself."

It is curious that Lagrange should carry all those dates in his mind; it shows that he had been thinking often about mathematical tradition. It must have touched young Poisson to be thus included by the old master in the society of immortals.

XVI. HIS VIEWS ON MATHEMATICS

Lagrange marked the end of the period initiated by Newton and Leibniz, he went as far as it was possible to do along the paths which they had opened, and his study of the old isoperimetric problems enabled him to discover a new path—the calculus of variations. One might expect a man who spent sixty years of his life meditating on mathematical subjects and on little else to have a great deal to say about mathematics, but that is not the case. After all that expectation is not a wise one; do the great musicians speak of the beauty of music? It is enough for them to create it, to talk about it would seem to them so much balderdash. However, Lagrange was perhaps a little more taciturn than most mathematicians.

It has been said by Lacroix who examined Lagrange's manuscripts (XII: 388) "that Lagrange did not neglect any mathematical publication of any importance and studied it pen in hand." This is probably an exaggeration, yet one cannot help envying the geometers of the eighteenth century who were still able to review a considerable part of mathematical literature without losing sight of their own problems.

We have already remarked that as late as 1811 Lagrange found it necessary to explain that the new calculus was really as safe as geometry. This shows that the possibility of doubt had occurred to his mind, and no wonder when we consider the theoretical basis of the calculus in his time. The wonder is that so much could be done with so uncertain an instrument.[78] However great a man's genius may be it is limited; he is able to see ahead of his fellowmen but only to a certain point. Lagrange could not possibly forestall the work of Gauss and of Cauchy. By the way, he had read the first part of Gauss' *Disquisitiones arithmeticae* and expressed his admiration to his young colleague (XIV: 298–300). He was never in touch with Cauchy, who was only

[75] That second part (III–IV, 1802) edited by Lalande was very imperfect indeed (*Osiris* 1: 550–553, 1936).

[76] Reprinted in Bertrand's edition of the *Mécanique analytique* II: 318–326, 1855 and in the *Oeuvres* IV: 439–451.

[77] *Oeuvres d'Arago* II: 671, 1854.

[78] See the remarks of Emile Picard, "On the development of mathematical analysis," *International Congress of Arts and Science of St. Louis* I: 503, Houghton, Mifflin, 1904, concluding wisely, "The demand for rigor in mathematics has had its successive approximations, and in this regard our sciences have not the absolute character so many people attribute to them."

twenty-four at the time of Lagrange's death. The introduction of more rigorous methods of analysis was very largely Cauchy's task.

Lagrange did not like the mixture of geometrical and analytical methods. Maurice remarks apropos of this (1819: 172),

> He believed firmly in the superiority of analysis, if he believed firmly in anything, for that extraordinary man was naturally skeptical and his own judgments were always expressed in a doubtful form, except when contradiction raised in him a tenacity out of harmony with his character. One has seen him extend that distrust of his opinions to the very results of his computations, and remain somewhat astonished that an exact deduction from the best lunar observations bore out completely a beautiful law which he had discovered concerning the movements of libration of our satellite. Indeed, with all the penetration of his mind, that man who knew so many things was afraid of all that one did not know.

Nothing illustrates better the limitation of genius, his inability to anticipate the future, than the following statement made by Lagrange in a letter to d'Alembert (Berlin, September 21, 1781; XIII: 368):

> Physics and chemistry are now offering riches more brilliant and easier to exploit; therefore, the reigning fashion is turned in their direction, it is not impossible that the seats reserved for Geometers in our Academies will be someday comparable to the chairs of Arabic in the Universities.

Think of the fabulous progress of geometry after Lagrange's death, which he could not possibly have imagined even in the faintest manner, without ceasing to be what he was,—a man, albeit great, of the eighteenth century.

Lagrange's greatness appears in his power of abstraction and generalization which enabled him to create the calculus of variations and after having given a suitable analytical form to the very simple principle of virtual velocities to derive from it the whole of mechanics. He was exceedingly fastidious with regard to the mathematical form of his writings. In 1769, that is in his youth, he wrote to d'Alembert (XIII: 132) "I have a bad habit which I am unable to shake off: I often rewrite my articles, even many times, until I be passably satisfied with them." It was not a bad habit and he did not try hard to shake it off; on the contrary the habit grew on him.[79] He was a consummate artist and his standard of mathematical elegance was exceptionally high.

XVII. LAGRANGE WEATHERS THE FRENCH REVOLUTION. HIS LAST ACTIVITIES

Lagrange left Berlin in 1787, because he was afraid that during the period of reaction which had set in after the death of Frederick the Great his peace of mind would be jeopardized. Disregarding Italian invitations he accepted the Parisian one. That was natural enough, for his best friends or sympathizers were in Paris, and that city had always attracted him. Nowhere did better conditions obtain for his meditations. Lagrange's choice was prudent, but he was not a prophet and did not begin to imagine to what kind of new perils his peace of mind would soon be exposed.

It is easy for historians writing a long time after the events to explain how revolutions are slowly prepared. Yet the preparation is largely unconscious and secret, and contemporaries, even those who are placed in the best observation posts, do not see the dangers menacing them. This was certainly the case for the French Revolution. Discontent grew steadily from 1715 to 1789, more rapidly after 1750, more rapidly again after 1770, yet nobody visualized a revolution, which happened as a surprize when it finally happened.[80] Lagrange gives us an additional proof of this, for well informed and prudent as he was if he could have foreseen the events of 1789, he would certainly not have established himself in Paris in 1787.

Lagrange's attitude to political problems has already been outlined apropos of his commentary on Denina's misfortune. His main concern was to preserve his peace of mind and keep out of trouble. In addition political problems must have discouraged him because of their inherent and excessive indetermination.

Upon his arrival in Paris he had been welcomed by the Queen and given lodgings in the Louvre. Proofreading of the *Mécanique* fully occupied his mind. He was apparently happy until the Revolution, though increasingly more anxious as the clouds gathered. July 14, 1789, was frightening, but there was reason for hope and a great burst of enthusiasm. The National Assembly (Constituante) confirmed Lagrange's pension and the latter continued his meditations. He seemed to

[79] Witness the letter to Father di Caluso (1784) quoted in chapter VII.

[80] Daniel Mornet, *Les origines intellectuelles de la Révolution française*, 1715–1787, Paris, 1933; *Isis* **23**: 463–467.

feel reassured and wrote to an unidentified prince (October 24, 1791; XIV: 283),

> Now that tranquillity and order are reestablished, I do not regret having attended a spectacle, the most interesting to philosophers, that of a great nation creating a new government not by the force of arms, but that of speech and public opinion.

When his pension became insufficient because of the depreciation of money, he was appointed a member of a new committee to reward useful inventions and later (March 1792; Maurice, 1819; 163) one of the three administrators of the Mint. There was a fine precedent for such a position; had Newton not been Warden then Master of the English Mint and had he not kept that office until the end of his life (1695, 1699–1727)? However, Lagrange was less patient with administrative details and remained only half a year in office. Moreover, the comparison is misleading. Newton had been given a high position, of great responsibility and steadiness, and very well paid. On the contrary, Lagrange was invited to take part in hopeless deliberations during a period of financial chaos. His job was as unpleasant as Newton's was comfortable. In the meanwhile, he married Mlle. Le Monnier (May 31, 1792) and anticipated a new lease of happiness. It is clear to me that the marriages of the two sisters Le Monnier must be interpreted in their revolutionary context; it was a time when the need of personal security was so great that many things were done for the sake of it which probably would not have been done in other circumstances. Unfortunately, the more ugly part of the Revolution began a few months after Lagrange's marriage with the opening of the National Convention on September 21, and it lasted as long as the Convention itself, three endless years. In 1793, Lagrange was for a moment in danger of being expulsed as an enemy alien, but Lavoisier without consulting him wrote a noble letter to Lakanal (September 7, 1793; XIV: 314–315). Alas! this might have done Lagrange more harm than good, but happily for him another chemist, Guyton de Morveau (1737–1816), whose revolutionary purity was beyond suspicion, came to his rescue and caused him to be requisitioned for ballistic computations. Lagrange was saved.[81]

To return to Lavoisier, we don't know how intimate was Lagrange's relationship to him, but it is certain that Lagrange admired him very much. He could appreciate the full value of the "chemical revolution" and especially the stupendous fact that chemistry was finally put on a clear quantitative basis. He studied the new science which had now become "as easy as algebra" (I: xxxviii). Lavoisier defended Lagrange on September 7, 1793, but soon afterwards the roles had to be reversed. Being a farmer-general and the son-in-law of another, Lavoisier was particularly obnoxious to the revolutionaries. On September 10–11 his house was searched for incriminating documents; on November 28 he was made prisoner at Port-Royal (then called Port Libre; in Paris). The newly created (temporary) Commission of weights and measures asked for his release on December 18, and as a result was itself "purged" on December 23 not only Lavoisier, but Borda, Laplace, Coulomb, Delambre, Brisson, Cassini IV, Méchain being removed; strangely enough Lagrange was kept.[82] The Advisory board of arts and trades, presided over by Lagrange, sent to Lavoisier a public certificate of its esteem.[83] On May 8, 1794, Lavoisier was condemned to death and executed on the same day. On the morrow Lagrange said to Delambre "It has taken them only a moment to cause that head to fall, and a hundred years may not suffice to produce a similar one" (I: xl).

By this time we can easily imagine that there was not much peace of mind left to him. The best that people like him could do during the Terror was to become as obscure, invisible, inaudible as possible and do their best to be forgotten. It was probably then that he took the habit of prefacing most of his answers with the words "*Je ne sais pas. . . .*" (I do not know. . . .) Those words are often quoted as a proof of his modesty; I prefer to take them as a proof of his prudence. There was no end of know-it-all

[81] The same Guyton de Morveau might have saved Lavoisier but would not. Maurice (1819: 164). Georges Bouchard, *Guyton-Morveau*, 335, Paris, 1938; *Isis* **32**, 153–156.

[82] He himself said, "I don't know why they kept me." The reason I suspect was that the others had probably compromised themselves in semi-public or private conversations, while he had kept his mouth shut. The remaining members of the purged commission were Lagrange, Berthollet, Haüy, Monge, Vandermonde (Bigourdan, 1901: 53).

[83] M. Berthelot, *La révolution chimique*, 201, Paris, 1890. Berthelot does not give the date of that certificate; if it was posterior to the declaration made by the Commission of weights and measures, Lagrange showed a courage which was unusual among men of science of that time. Lavoisier was abandoned by most of them; it must be recognized that his case was hopeless.

people during the Revolution, and many of them were so sure of their knowledge that they did not hesitate to condemn their contradictors to death. Never does the danger of ignorant dogmatism appear more blatantly. Lavoisier's answer "I don't know" was thus natural enough; it was a gentle reproof at once to the idiots who credited him with knowledge he could not possess and to the know-alls who had an answer to everything, a bloody one if no other was handy.

We do not know exactly what happened to Lagrange. He thought for a moment of leaving, even of returning to Berlin. The details of his perplexities escape us; however great his fears may have been, he finally decided to remain in Paris, partly because of his Parisian wife, partly we may assume, because they could not have left with their property without running great dangers. Alas! it is easier now to imagine their bewilderment, because so many of our own friends have experienced it recently. The beasts of the French Revolution were savage enough, yet they were on the whole less cruel and more excusable than the Nazi ones.

The Metrical System was promulgated by the Convention on April 7, 1795, and the twelve following commissaries were appointed on April 17, for the discussion and execution of all the scientific operations: Berthollet, Borda, Brisson, Coulomb, Delambre, Haüy, Lagrange, Laplace, Méchain, Monge, Prony, Vandermonde.[84] Lagrange played an essential part in defending the integrity of the decimal idea.

When the Bureau des Longitudes was created, upon Lakanal's proposition, on June 25, 1795, Lagrange was one of the original geometers, the other being Laplace. The other foundation members were the astronomers Lalande, Cassini IV, Méchain, Delambre, the ancient navigators Borda and Bougainville, the geographer Buache, and the artist (meaning instrument maker) Carochez. The metrical system having been finally sanctioned on December 10, 1799, the First Consul decreed on September 24, 1803, that the standards would be preserved by the Bureau des Longitudes. The Bureau then became the guardian of the metrical system.

The elder Mirabeau, whom we have already met in Berlin, had written early in 1791 a treatise on public education[85] wherein he recommended that the academies should be controlled by the nation, instead of being royal foundations. On August 8, 1793, the Convention suppressed the academies, not excepting the Academy of Sciences, in spite of the latter's obvious usefulness. This was natural enough for the elections were controlled by the king, the salaries of the regular members were "royal pensions,"[86] and some members were appointed because of their nobility rather than their merit. At Lakanal's request, an exception was soon made for the Academy of Sciences yet it remained closed. Mirabeau's ideas of 1791 were realized only on October 25, 1795, when the Institut National was created; the old *academies* became *classes* of the Institut.[87] The first class devoted to physical and mathematical science held its first meeting on December 27, 1795. The original members appointed by the Directoire had been most of them members of the Royal Academy. The first officers of that class were Lagrange, President; Laplace, Vice-president; Lacépède, Secretary and Haüy, Vice-secretary (Gauja, 1934; 28; *Inst. de France, Procès-verbaux* **1**, 1910).

Lagrange did not care for teaching and had refused in 1787 a generous invitation of the University of Turin, but in revolutionary times one had to do many things one did not choose. It was of supreme importance for a man situated as Lagrange was, to prove his usefulness. He was one of the first professors in the short-lived Ecole Normale of the year III (1795),[88] teaching arithmetic, algebra, and their applications to geometry. The "elementary lessons" which he gave at the Ecole Normale were published immediately in the *Séances des Ecoles normales* (an III, 1794–1795); seventeen years later Lagrange had them reprinted in the *Journal de l'Ecole Polytechnique* (cahiers 7, 8, 1812). *Oeuvres* VII:

[84] That is the old personnel, except Cassini IV and of course Lavoisier.

[85] Mirabeau died on April 2, 1791; his treatise was published posthumously the same year by P. J. G. Cabanis, *Travail sur l'éducation publique trouvé dans les papiers de Mirabeau l'ainé*, Paris, Imprimerie Nationale, 1791.

[86] Therefore, the regular members were called "pensionnaires." When Lagrange came to Paris, his position was that of "vétéran pensionnaire" and his salary was paid like the others out of the privy purse.

[87] The word académie was dropped as being too suggestive of the old régime; the classes were renamed académies only in 1816. In the present organization not essentially different from that of 1816 the Institut de France includes five academies, and thus old and new ideas have been brought together, the academies being originally royal foundations, the institute a revolutionary one. The five academies celebrate each year together the birthday of the Institut, October 25.

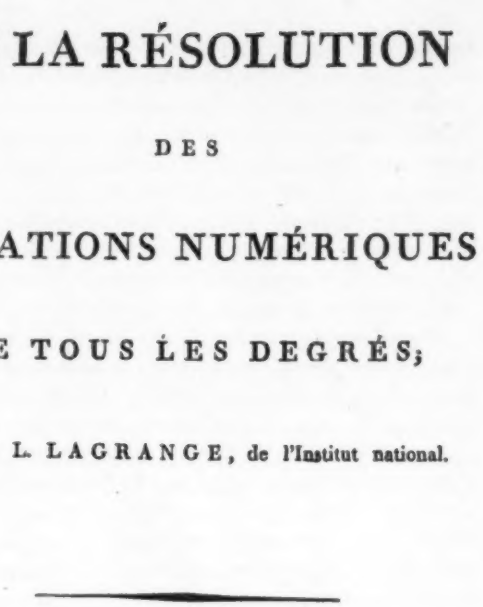
[88] The Ecole Normale existing today was created only in 1810.

183–288. An English translation of those lessons; *Lectures on elementary mathematics*, was made by Thomas J. McCormack, Chicago, Open Court, 1898. The "Essai d'analyse numérique sur la transformation des fractions" published in the same *Journal*, cahier 5, **2**, May, 1798 (*Oeuvres* VII: 291–313), is probably another fruit of his teaching experience of that time.

He was also one of the earliest teachers of the Ecole Polytechnique, and he is thus one of the founders of the two greatest schools of France. When the "revolutionary courses" began in the Ecole Polytechnique on December 21, 1794, the teaching of analysis and mechanics was intrusted to him and Prony.[89] The ordinary courses opened on May 24, 1795, the first being delivered by Lagrange in the presence of the whole school, students and faculty. In 1799, he resigned because of old age (he was sixty-three) and was succeeded by Lacroix; he remained a member of the "conseil de perfectionnement" of the school, being one of those delegated by the Academy of Sciences (the other academic delegates were Laplace, Monge, Berthollet); in 1813 he was replaced by Lazare Carnot.

According to Polytechnical traditions Lagrange was a poor teacher because he was soaring too high.[90] That is plausible, though it would be foolish to assume *a priori* that he could not explain himself because his thoughts were too deep; depth and simplicity may come together as well as not. The lessons he gave at the Ecole Normale were very simple enough. When he was obliged to resume teaching in 1795, he was almost sixty and had been out of practice for forty years. He was too indrawn, more adept in following steadily and quietly his own ideas than in educating young men. His heart was not in teaching, and therefore it is probable that he was not a good teacher.

Three didactic treatises were published by Lagrange during the Directory, the final period of the Revolution, the *Théorie des fonctions analytiques* in 1797,[91] the *Résolution des équations numériques* in 1798, the *Leçons sur le calcul des fonctions* in 1799. I do not know how closely the first two were related to his teaching, but the



DE LA RÉSOLUTION

DES

ÉQUATIONS NUMÉRIQUES

DE TOUS LES DEGRÉS;

Par J. L. LAGRANGE, de l'Institut national.

A PARIS,

Chez DUPRAT, Libraire pour les Mathématiques, quai des Augustins.

AN VI.

FIG. 13. First edition, Paris, 1798.

third reproduces lessons actually given at the Ecole Polytechnique in 1799. Thus we may say that that book was published in 1799, though it was printed only in 1801 (in *Séances des Ecoles Normales* **10**). An introductory discourse was published in the *Journal de l'Ecole polytechnique* **2**, cahier 6, July, 1799, *Oeuvres* VII: 325–328.

During the Empire, that is, during the last period of Lagrange's life, he occupied himself in revising and amplifying his former works. The *Leçons* were reprinted in the *Journal de l'Ecole Polytechnique* (cahier 12, 1804); revised edition with two more lessons in 1806; these two additional lessons were also published in the *Journal de l'Ecole Polytechnique*, cahier 14, 1808; *Oeuvres* X, 1884.

A revised edition of the *Résolution des equations numériques* was printed in 1808.[92] That edition

[89] Baron de Prony (1755–1839), mathematician and engineer after whom the Prony brake is named.

[90] *Livre du centenaire de l'Ecole Polytechnique* I: 19–21, 25, 40, 47, 1895.

[91] Lagrange presented it to the Institut on June 5, 1797 (*Inst. de France, Procès-verbaux* **1**: 219, 1910). The title page (see facsimile) bears the date Prairial an V = May–June 1797.

[92] Lagrange presented it to the Institut on March 14, 1808 (*Inst. de France, Procès-verbaux* **4**: 29).

was reprinted in 1826 and in the *Oeuvres* VIII, 1879.

A new edition of the *Mécanique analytique* began to appear in 1811, the second volume of it appearing only posthumously in 1815 (see previous section).

As to the *Théorie des fonctions analytiques*, as soon as it was published (1797), it was closely examined by a German mathematician, Johann Philipp Grüson (1768–1857), who prepared a German translation of it and sent the author a list of eighty-five corrections and a long letter dated July 14, 1798 (XIV: 289–293). Grüson's translation was completed in the same year and published in two volumes (Berlin, 1798–1799), together with a note of Lagrange on the principle of virtual speeds (II: 266–272).[93]

The *Théorie* was translated also into Portuguese under the title *Theorica das funções analyticas* (2 v., Lisboa, 1798). That translation may be called a Brazilian achievement for its author was Manoel Jacinto Nogueira da Gama, born at S. João de El-Rey, Minas Gerais, September 8, 1765, educated in Coimbra, returned to his native country and took an important part in the colonial administration, also as a soldier, sailor, and statesman. He served in both the Portuguese Navy and Army and retired with the rank of marshall. He held in Brazil several political and administrative positions: deputy to the Constituinte, senator from the province of Minas Gerais, president of the senate, and minister of state. He was created first marquês de Baependi. After the abdication of the first emperor, D. Pedro I, on April 7, 1831, he retired from political life; he died in Rio de Janeiro on February 15, 1847.[94] He translated into Portuguese not only Lagrange's *Théorie*, but also the *Réflexions sur la métaphysique du calcul infinitésimal* published by Lazare Carnot in 1797.[95] His own publications were devoted to practical questions of hydraulics, chemistry, etc., which were of immediate importance for the development and defense of his country.

A revised and augmented edition of the French text was published in Paris at the beginning of 1813; Lagrange presented it to the Institut on February 22 (*Inst. de France, Procès-verbaux* **5**: 168), that is, at one of the last meetings which he could attend (he died on April 10). A third edition was issued in 1847 by Alfred Serret who added a note to it. A fourth edition, with the same note, was included by the same editor in the *Oeuvres* IX, 1881.

Bonaparte, remembering the mathematical studies of his own youth, was ready enough to appreciate and reward mathematicians. The story of the delegation sent to Lagrange's father to congratulate him in 1797 has been told in chapter III. During the Italian campaign Bonaparte had become acquainted with another mathematician, Lorenzo Mascheroni, who had dedicated to him ("A Bonaparte italico") *La geometria del compasso* (Pavia, 1797). When Bonaparte came back from Italy he amused himself with teasing Lagrange and Laplace and asking them to solve simple geometrical problems, according to Mascheroni's method, that is, with compass alone (*Isis* **26**: 168). One of these problems "How to determine the center of a given circle with compass alone" has even been called "Napoleon's problem." It is a good puzzle. Try it!

On the other hand, we may assume that Lagrange admired the First Consul and later the Emperor. Were not peaceful men, above all men of science, grateful to the savior who had brought back order and tranquility after the prolonged revolutionary nightmare? At first Napoleon gave his old followers the impression that he remained faithful to the main purpose of the Revolution; after having put an end to the excesses he wanted to preserve what was best in the revolutionary ideals. It is probable that men like Lagrange did not realize the perversion of those ideals implied in the Napoleonic adventures. And what did he think of the turn of events during the Peninsular War or at the time of the disastrous retreat from Russia? If his wife or others questioned him he probably replied "I do not know . . . ," and that was true enough. How could he know? How could he foresee the battle of Leipzig, the entry of the allies into Paris, the Hundred Days, St. Helena? How could he have guessed that the Napoleonic "new order" was but an empty shell? We may count him happy to have died before the final disasters occurred; it was enough for him to have witnessed one revolution.

Napoleon was very considerate of Lagrange,

---

[93] That note was published in French in the *Journal de l'Ecole Polytechnique*, cahier 5, May 1798; reprinted in *Oeuvres* VII: 317–321.

[94] Information kindly communicated to me by Senhor Rodolfo Garcia, Diretor Geral, Bibliotéca Nacional, Rio de Janeiro (letter of March 1, 1944).

[95] *Reflexões sobre a metaphysica do calculo infinitesimal*, Lisboa, 1798.

but it is probable that Lagrange treated his new sovereign as he had treated Frederick the Great, that is, with the utmost respect but no desire of wasting his own time. In a case like this respect became the best defense of his personality, one could show the sovereign so much respect that no familiarity was possible. A man like Lagrange had no time to be a courtier. Napoleon made him a senator and a count. These distinctions did not honor Lagrange, nor yet dishonor him. They may have seemed very important, even to himself, at the time he received them. However, his fame is independent of such ephemeral futilities, it is built on a more solid foundation, the *Mécanique analytique* and other monuments *aere perenniora*.

XVIII. LAGRANGE'S TEMPERAMENT, HEALTH, AND DEATH

The best information concerning Lagrange's physical life is given in the *Précis historique* written soon after his death by Julien Joseph Virey, who was one of his physicians during the last period of his life and obtained pertinent data from Lagrange himself and from the comtesse Lagrange.

Lagrange was of middle height and in his youth was thin and pale (*grêle, mince, et pâle*). As soon as his mathematical vocation had declared itself, that is, before he was eighteen, he assumed the habit of working with great concentration. Most of his work was purely mental, and he was generally pensive and silent. Sometimes his thinking would throw him into a kind of feverish excitement, and his pulse would become irregular. He had pulse irregularities until the end of his life. At about the age of twenty-five at the end of the winter he suffered a hypochondrial[96] affection (I am using of necessity Virey's terms) and displayed bilious symptoms. He became definitely nervous and melancholic. The doctors of Turin and Berlin bled him repeatedly; during the course of his life, his blood was let twenty-nine times. He was very careful as to diet and regimen, and his interest in such matters caused him to read medical books, to study the nature of poisons and other dangerous substances, and to form physiological opinions of his own. For example, he fancied that the complexion which the ancient called melancholic (that was his own complexion) was caused by the predominance of the venous over the arterial system.

Lagrange managed to reestablish his physiological equilibrium by an interruption of his studies and a restorative diet. He used lemon juice together with fruits and greens. His favorite diet was always vegetarian, "almost Pythagoraean." Virey assures that this contributed not a little "to the moderation of Lagrange's character and the peacefulness of his soul."

The climate of Berlin was apparently more favorable to him than that of Turin. He had never drunk wine, except diluted with water and in small amount; in Berlin he cut it out completely, replaced it with beer and felt the better for it. He became a little stouter.

His sedentary and studious life exposed him to hemorrhoids and to a kind of overheating which he tried to correct with vegetal acids, such as lemon juice in tea. At the end of each winter he suffered from bilious defluxions and spells of hypochondria. We can guess a few of his troubles from his correspondence with d'Alembert. On September 21, 1767, the latter ascribed them to overwork, lack of recreation, excessive use of tea and coffee and advised him to read Tissot's recent book on the diseases of men of letters[97] (XIII: 100). On June 17, 1770, Lagrange complained of a fever which had kept him down for many days (XIII: 173). D'Alembert answered "I notice with fear and sadness that you fall gravely ill each year, your illness being certainly caused by excessive work. Take good care of yourself, I beseech you . . ." (XIII: 178). D'Alembert gave him good advice but bad examples, for he was less able to husband his strength than Lagrange was; he advised travelling but neither of them managed to steal enough time from his work for such a purpose.[98]

It would have been better perhaps for Lagrange to relax his mind more frequently as d'Alembert suggested, but his physical hygiene seems to have been on the whole a wise one. Born with a delicate constitution he built up his strength by means of regular walks and a very moderate diet. He had taken from Frederick II

[96] "Hypochondria" is a kind of depression or "morbid anxiety as to one's own health, with conjuring up of imaginary ailments" (Webster).

[97] D'Alembert was referring to the famous Swiss (Vaudois) physician, Auguste Tissot (1728–1797), practising in Lausanne. The book wrongly quoted by him was Tissot's *De valetudine litteratorum* (Lausanne, 1766), which Tissot translated himself into French, *De la santé des gens de lettres* (Lausanne, 1768). Often reprinted in many languages.

[98] D'Alembert did sometimes travel but was always very anxious to be back in Paris as soon as possible.

the habit of extreme regularity, that is, of doing always the same things at the same times. He found that work thus regulated became easier and more agreeable, and in spite of his native industriousness, he never ceased giving himself each day a definite task for the following.[99] "The spirit is lazy," said he, "one must always be on one's guard to prevent its natural slackness and to develop its strength for future needs" (Maurice, 1819: 172).

At the age of forty-two (in 1778) he had a bilious peripneumonia and suffered from a dysuria caused by the use of cantharides (Virey, 1813: 9).

Virey describes Lagrange as he knew him in Paris. Middle height, delicate complexion, a calm and pensive look, his glances penetrating but gentle. His hair which had been brown was white but plentiful. He was timid and gentle and spoke slowly, rolling a little his *r*'s, lifting his voice when he was discussing a matter of interest to him. He could then find without difficulties the words needed to express his ideas clearly and strongly. He did not like rambling conversations and even less idle ones. He was earnest without sadness, and always of sweet temper. Equitableness was to him like a mathematical rule, which one cannot violate without absurdity. Fairness and kindness were the principles of his conduct.

His stomach was poor and easily upset; therefore, he did not like to dine out and his ordinary diet was abstemious. His memory, intelligence, and capacity for work remained good until the end. His sight remained good too, but he became hard of hearing.

He got up late (about ten o'clock), and took coffee with milk; he dined soberly eating but little meat, mostly vegetables and fruits. He drank beer then light coffee, very sweet, and swallowed it as hot as possible. At ten o'clock in the evening he drank tea, very hot and sweet, sometimes with lemon, and went to bed about midnight.

He loved to walk in the open air, and to be alone when he did so for walking helped him to meditate.

For the long period 1795–1813, the minutes of the mathematical class of the Institut record only once his absence from a meeting because of illness; that was on January 31, 1803 (*Inst. de France, Procès-verbaux* **2**: 625).[100]

Toward the end of his life he suffered various fainting fits; a more severe one than usual occurred in February 1813. His wife found him unconscious in his study. In falling against the side of a table he had hurt the left side of his head. He recovered but had a bad cold on March 28, was feverish and out of sorts. He decided on his own accord to take a grain of emetic, and vomited a considerable amount of bile. After that he fainted many times, and had pains in his arms and legs; his skin became so sensitive that he could not bear the slightest contact or the weight of his bed-clothes.

On April 2, he consulted Dr. Potel and later Dr. Virey. The latter describes his illness with greater detail than is desirable here. On April 8, his colleagues the senators counts Monge, Lacépède, and Chaptal visited him and brought him a new imperial decoration, he spoke with them with some animation for more than two hours.

> "I have been very ill the day before yesterday my friends," he said, "I felt that I was dying; my body was weakening little by little, my faculties moral and physical were imperceptibly fading away; I was observing with pleasure the gradual diminution of my strength and was reaching my goal without pain or regrets along a soft declivity. Oh! death is not to be feared and when it comes without pain it is a last 'function' which is neither distressing nor disagreeable." . . . "I wanted to die," he added with greater force "yes, I wanted to die, but my wife did not want it. At that moment I would have preferred a wife who was not as good, was less anxious to revive my strength, and would have left me to end softly. I have had a long career; I have obtained some fame in mathematics. I have never hated anybody, I have done no harm, and there must be an end, but my wife did not want it" (I: xliv).

Then he proceeded to tell the history of his life, of his works, of his days in Berlin and Paris, etc. During that long conversation, he had sometimes difficulty in remembering names and dates, but did not ramble and was able to express vigorous thoughts in good order and with bold images.

That last conversation of his to be recorded is very moving. One wonders what caused the deep lassitude which it betrayed? Surely not

[99] One should bear in mind that he was his own taskmaster; he had no chief, no assistant, no student, but was continually challenged by the publications of his few rivals.

[100] It would be more prudent to say I noticed only one such record; there may have been others (very few) which I failed to notice. Or other absences may have remained unrecorded.

his work, intense as it had been, for that had been the main solace of his life, his "raison d'être" to which everything else had been subordinated. Until the Revolution he had been extremely privileged, for he could do his own work as he chose with practically no outside obligations. Not only was he freed from teaching, but unlike the senior members of our research institutes, he had no administrative worries of any kind. He worked entirely alone, and his only external obligation was to attend regularly academic meetings; these meetings were simple as compared with those of our own academies, and the business or administrative part of them reduced to a minimum. Therefore, it is probable that his lassitude had a different cause. The Revolution must have been a terrific strain on him and the more so because he had to remain silent. He was a kind man, and though he did not speak of it the brutalities and atrocities of the Terror must have wounded him deeply. It is easy for any sensitive man of our own time to understand that, for is he not distracted almost beyond endurance by the atrocities which are reported in every day's papers? And yet those atrocities happen far away, while they were very close to Lagrange. Some of the scientific committees of which he was a member were "purged" and the academy itself was closed by the Convention on August 8, 1793. Toward the end of 1795 one could breathe more easily again, but Lagrange was now charged with many duties and getting older. During the Consulate and the Empire many honors were showered upon him, but it is probable that he never recovered the tranquillity of earlier years. During the last years of his life (1812–1813) the Empire was beginning to totter. Would it be necessary to face another cataclysm? No wonder Lagrange was tired and willing to go. Life had been long enough. He had done all he could; what was the point of weathering new storms without profit to anybody? He was tired to death.

The valedictory meeting with his colleagues had exhausted him, and the following night was pretty bad. On April 9, he was completely prostrated. Dr. Portal was called in consultation. At the end of that day, Lagrange was still conscious. After another bad night he died on April 10, 1813, at nine-forty-five in the morning in his own house (rue du Faubourg Saint-Honoré no. 128) at the age of seventy-seven years, two months, and ten days. The original death certificate was destroyed in 1871 during the Commune, but a copy of it was preserved.[101]

On April 11, a post mortem examination was made by Dr. Potel, assisted by Boudet, pharmacist and embalmer. The official report of the autopsy is given by Virey (1813: 21–22). We need not discuss it. Lagrange's life had been a very long and active one; as he himself realized it in extremis, he had reached without excessive jolts its natural term.

He had studied mathematics for sixty years with hardly any interruption and had been at the front rank of mathematicians for half a century.

Solemn obsequies were arranged by the Conservative Senate; they took place in the Panthéon on Tuesday April 13, 1813, at noon (*Inst. de France, Procès-verbaux* **5**: 197). Speeches were delivered in the name of the Institut by Laplace and by the naturalist Lacépède.[102] Lagrange is buried in the Panthéon.

APPENDICES

XIX. LAGRANGE'S COLLECTED WORKS

Oeuvres de Lagrange, 14 v., Paris, 1867–1892, edited (I–X) by Alfred Serret (1819–1885) and after his death (XI–XII) by Gaston Darboux (1842–1917).[103] That edition is very convenient, especially because Lagrange's original memoirs distributed in the publications of the academies of Turin, Berlin, and Paris, are available only in very large libraries, and even there are not always easy to find. It is generally praised, with a kind of superstitious confidence. It is very unsatisfactory to the historian. The editor followed the path of least resistance putting together first the Turin memoirs, then the Berlin ones, then the Paris ones, then the rest. Corrections were almost exclusively restricted to mathematical ones; there are hardly any historical notes. The editor had no historical training, no idea whatsoever of the responsibilities which the publication of ancient texts implies.

Such an undertaking should always begin with

---

[101] Text given by Loria, *Scritti*, 331, 1937. Lagrange is described in that official document as "sénateur, grand officier de la Légion d'honneur, membre de l'Institut et du Bureau des Longitudes."

[102] Laplace's speech is not included in his *Oeuvres* (14 vols., Paris, 1878–1912), but Lagrange is quoted in almost every volume (see vol. 14, 460). Nor is the other speech included in the *Oeuvres du comte de Lacépède* (I, *Discours*, Paris, 1826).

[103] Volumes XIII–XIV will be discussed in chapter XXI below. The following remarks concern only volumes I–XII.

the printing of the correspondence in order that the letters might be conveniently used and referred to for the elucidation of the memoirs and books. This was done admirably in the case of Huygens, the edition of whose works[104] is a paragon. For the printed or unpublished memoirs and books one of two methods might be followed. They might be published either in strict chronological sequence, or divided into groups by subjects, the chronological sequence being adhered to in each group.[105] In any case it is not permissible to reprint a memoir first printed a century earlier without adding a number of notes, explaining the circumstances of the original publication and its repercussions. Of course, this requires considerable labor which Serret and Darboux were neither willing nor capable of doing.

XX. LAGRANGE MANUSCRIPTS

In the meeting of the Institut of May 22, 1815, the Minister of the Interior, Lazare Carnot, announced that the manuscripts left by Lagrange had been bought by imperial order, to be preserved in the library of the Institut (*Inst. de France, Procès-verbaux* **5**: 513; *Oeuvres* XIV: ix). A commission was appointed to examine them. It included Legendre, Prony, Poisson, and Lacroix; its report written by Lacroix[106] was deposited on November 3, 1817, and has been printed in the third edition of the *Mécanique* and in *Oeuvres* XII: 387–388. It concluded that none of the manuscripts should be published, but that they should be classified and bound in sixteen volumes. A few extracts were printed in the *Oeuvres*.

Thanks to the kindness of Emile Picard (1856–1941), Perpetual Secretary, Gino Loria was permitted to examine those manuscripts some time after 1913 in the library of the Institut, and he wrote to me on September 11, 1940, the following information (I translate freely from the French):

> The collection is bound in two series of volumes, some of quarto, others of octavo size. It contains much material for the edition of Diophantos[107] which Lagrange had projected, also on Euclid's postulate. Lagrange began a communication on the last named subject at the Institut and interrupted himself with the words "I must think of it longer."[108] It also contains a memoir ready to be printed on a theorem of Bernoulli relative to curves. A deeper investigation of those manuscripts would probably bring to light many other results deserving to be known. . . .

Loria's prediction is very likely to be correct. We should not forget that the members of the Institut to whom the examination of the manuscripts was intrusted were mathematicians, not historians, and they may have disregarded many manuscripts which added nothing to mathematical knowledge yet were or might be deeply interesting for humanistic reasons. It may be worth while to reexamine them very carefully after the end of the war.

XXI. CORRESPONDENCE

The bulk and the most valuable part of Lagrange's correspondence is included in the *Oeuvres* (XIII: 1882; XIV: 1892). That edition was intrusted to Ludovic Lalanne (1815–1898), a trained bibliographer and historian, though not a historian of science; it was done very conscientiously. Volume XIII is devoted to the letters exchanged with d'Alembert; volume XIV to those exchanged with Condorcet, Laplace, Euler and others. Volumes XIII–XIV are the only volumes of the *Oeuvres* which are indexed!

Other editions of the letters will be mentioned only in the case of letters not printed or reprinted in the *Oeuvres*.

(1) In the German translation of Lagrange's works, *Mathematische Werke*, undertaken by August Leopold Crelle (Berlin, 1823)—the same Crelle, who edited the famous *Journal für die reine und angewandte Mathematik* from 1825 to his death in 1855—are quoted in German translation two letters to Guyton de Morveau, one of which is dated March 2, 1797. After Bopp's paper of 1913, quoted below.

(2) Antonio Favaro. Sette lettere di Lagrange al P. Paolo Frisi, tratte dagli autografi nella Biblioteca Ambrosiana di Milano, *Atti R. Acc. Torino* **31**: 182–194, 1895.

---

[104] Vol. I appeared in 1888, vol. XX in 1940, not yet completed. See *Isis* **5**: 154–56; **21**: 213–15; **29**: 431–33; **34**: 51.

[105] The second method is followed in the Huygens edition and also in the Euler one undertaken in 1909 (*Isis* **1**: 244). Latest volumes mentioned in *Isis* **34**: 351. See also Pasquier, 1927: 116–119, When completed the Euler edition will include some sixty-nine large volumes!

[106] Signed by the four members of the commission and also by Frédéric Maurice.

[107] It would seem that Lagrange had really planned to prepare a new edition of Diophantos (Loria, *Scritti*, 1937: 310). He was deeply interested in Diophantine analysis which formed the subject of his additions to Euler's algebra (1774). There are various references to Diophantos in his correspondance with d'Alembert and Euler, 1768–1770. (G. S.)

[108] That story was told by Jean Baptiste Biot in *Journal des Savants*, February 1837, and was reprinted in his *Mélanges scientifiques et littéraires* II: 263, 1858. (G.S.)

Seven letters to the Barnabite mathematician Paolo Frisi (1728–1784), dated 1756 to 1782, all in Italian except the fourth one dated 1767 which is in French.

(3) PIETRO RICCARDI. Alcune lettere di Lagrange, di Laplace e di Lacroix dirette al matematico Pietro Paoli, *Mem. R. Acc. Modena*, ser. 3, **1**, sci.: 105–129, 1897.

A single letter from Lagrange to Pietro Paoli (1759–1839) dated September 1798, thanking him for his *Elementi d'algebra* (2 v., Pisa, 1794).

(4) *Catalogo della Biblioteca spettante all' eredità beneficiati del fu principe D. Baldassarre Boncompagni (1821–1894)*, Roma, 1898.

Including many Lagrange autographs listed by Bopp (1913, p. 46–47), see below. Some of these are edited in the *Oeuvres*.

(5) G. PITTARELLI. Due lettere inedite di Lagrange all' abate di Caluso esistente nell'Archivio storico municipale di Asti, *Atti del IV Congresso internazionale dei Matematici, Roma, 1908* III: 554–556, Roma, 1909.

Two letters dated Berlin, 1784, Paris, 1788, both in French, to Tommaso Valperga di Caluso dei Conti di Masino (1737–1815).

(6) FAGNANO. *Opere matematiche del marchese Giulio Carlo de'Toschi di Fagnano*, ed. by V. Volterra, G. Loria, D. Gambioli, III: 179–213, Milano, 1912; *Isis* **1**: 161, 1913.

Correspondence between Lagrange and Fagnano (1682–1766), twenty-one letters all in Italian. Fifteen dated 1754, four dated 1755–1756, two dated 1759. The contents are almost exclusively mathematical.

(7) ICILIO GUARECHI. Notizie storiche intorno a Luigi Lagrange, *Mem. R. Acc. Torino*, ser. 2, **64** (1), 1913.

Two French letters kept in the Municipal Library of Turin, undated (1791–1792?). One is addressed to Lagrange's brother "à Monsieur Lagrange. Receveur des hospices civils à Turin."

(8) KARL BOPP. Eine Schrift von Ensheim, Recherches sur les calculs différentiel et intégral [Paris, an VII, 1799] mit einem sich darauf beziehenden Brief von Lagrange. *Sitz. Ber. Heidelberg. Akad. Wiss., math. Kl.*, no. 7, 1913.

Letter to the citoyen Ensheim, dated November 23, 1799. Notes on other unpublished letter from or to Lagrange, p. 46–48.

Ensheim is otherwise unknown to me and the pamphlet analyzed by Bopp must be exceedingly rare. There is no copy of it in the Bibliothèque Nationale, Paris.

(9) CHARLOTTE H. BOATNER. Certain unpublished letters from French scientists of the revolutionary period taken from the files of Joseph Lakanal. *Osiris* **1**: 173–183, 1936.

JOSEPH LAKANAL (1762–1843), French educator and politician, left France after Waterloo and found refuge in the United States, where he spent many years (1816–1833). From 1817 to 1825 he was President of the Louisiana State University. His correspondence was long preserved in the Athénée Louisianais, New Orleans; it has recently been transferred to the Howard Library, which will soon be incorporated into the Tulane Library, New Orleans.

There is but one letter from Lagrange, addressed to Lakanal in November 1794. See text and facsimile reproduction of it in *Osiris* **1**: 178–179; it is a very fine specimen of Lagrange's writing, revolutionary style.

XXII. LETTERS AND OTHER AUTOGRAPHS IN THE UNITED STATES

Professor Boatner's article on the Lakanal correspondence gave me hopes that more Lagrange Letters might be found in this country. Indeed, Lakanal was not the only French refugee to come here; many had preceded him during the Revolution, others followed him after the Napoleonic period. Many of those refugees were very able men who exerted a marked influence upon American education and especially upon the teaching of mathematics.[109] I wrote to various libraries where such relics might be preserved, but the results of my inquiries are very meager. It is possible, however, that Lagrange autographs are still hidden in unexpected places, and this article may be the means of bringing them to light.

The only autographs thus far discovered in the United States are:

(1) The letter already referred to kept in New Orleans and published in *Osiris* **1**, 1936.

(2) A single letter in the Pierpont Morgan Library, New York, published and reproduced here with the kind permission of Miss Belle da Costa Greene, Director of that Library.

(3) Four documents bearing Lagrange's signature in the David Eugene Smith Library, Columbia University, New York. Another document (two sheets) in the same library does not concern the mathematician Lagrange but his contemporary, the general Lagrange.

(4) Photostat of a letter in Lagrange's hand in the David Smith Library, the location of the original being unknown. (It was probably a photostat submitted to Professor Smith by a dealer.)

(*A*) Undated letter in the Pierpont Morgan Library.

Nous aurons l'honneur ma femme et moi de nous rendre jeudi 13 à l'invitation de M. l'Ambassadeur et nous profiterons avec reconnoissance de l'offre obligeante de sa voiture.

ce jeudi 6 LAGRANGE

[No year or month is indicated, but the letter is posterior to his second marriage on May 31, 1792. In the dating of it samedi was changed into jeudi. This letter formed part of a collection of miscellaneous autograph letters purchased by Mr. Morgan before Miss Greene came to the library, and there is no record of the source of acquisition.]

---

[109] Lao Genevra Simons, The influence of French mathematicians at the end of the eighteenth century upon the teaching of mathematics in American colleges, *Isis* **15**, 104–123, 1931. Reprinted in her book *Fabre and mathematics*, N. Y., Scripta Mathematica, 1939.

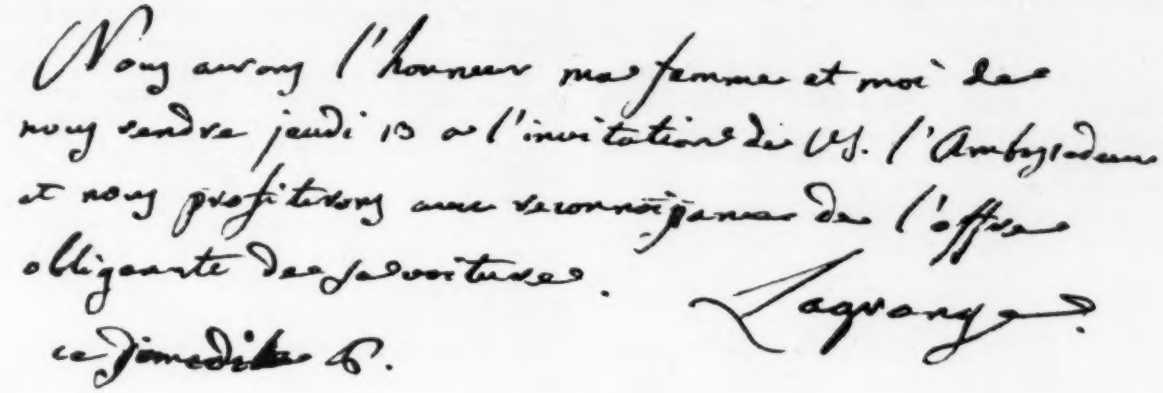


FIG. 14. Autograph *A*. Note written by Lagrange in or after 1792 (Pierpont Morgan Library).

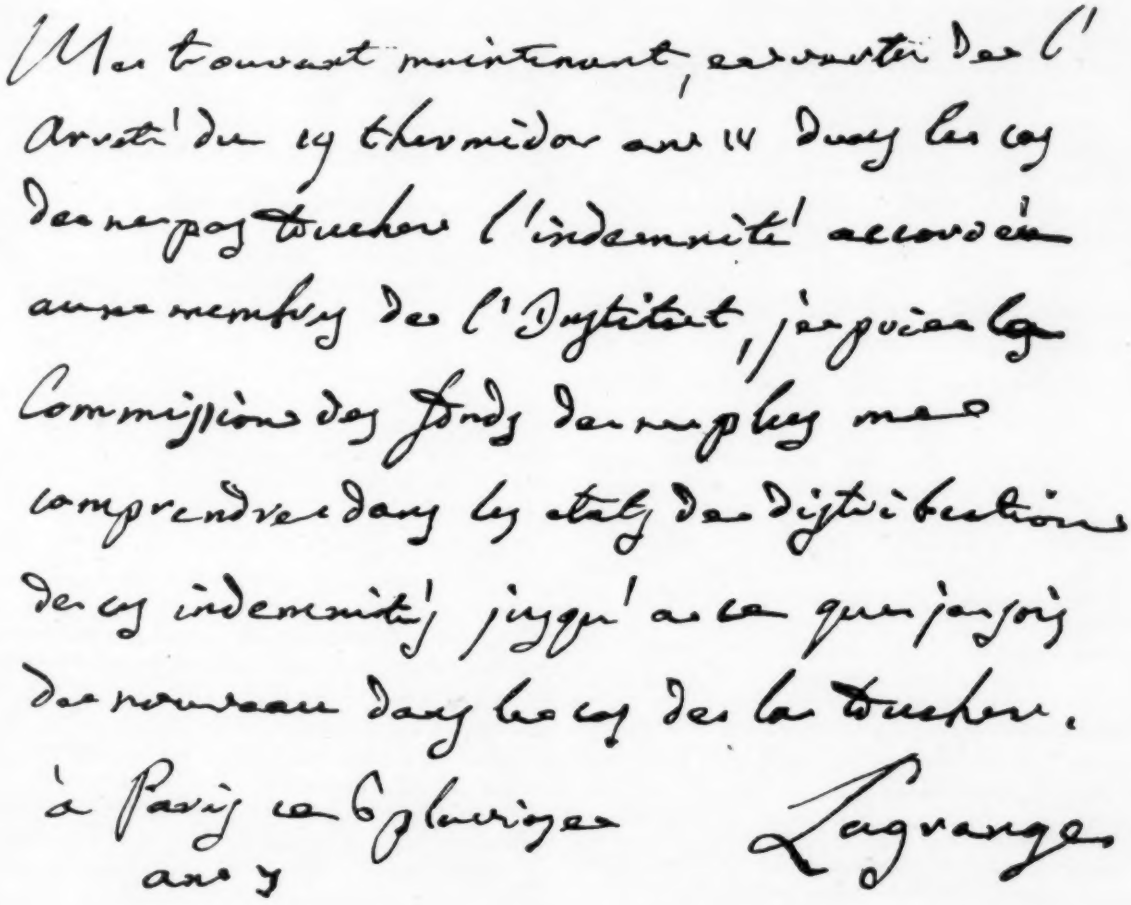


FIG. 15. Autograph *D*. Letter written by Lagrange in 1797 (present location unknown).

(*B–F*) Documents in the David Eugene Smith Library.

(*B*) Nous Commissaires Généraux des Monnoies autorisons le Sieur Savari préposé à la recette des matières destinées à la fabrication de la menue monnoie dans l'atelier des Barnabites à remettre à M. Brun douze livres de métal de cloches.
Paris le 9 juillet 1792, L'an 4e de la Liberté.[110]

Signed by seven men including MONGE, BERTHOLLET and LAGRANGE

(*C*)

Paris le 11(?) Nov. 1792
L'an 1er de la République
Citoyen

La Convention Nationale nous demande Messieurs, le montant des dépenses de l'hôtel des Monnoies de Lyon,† En conséquence nous

†Pour l'année 1793

vous prions en exécution de la Loi du 27 mai 1791 de vous réunir à nous pour pouvoir présenter à la Convention Nationale les prix auxquels s'éléveront les réparations à faire au dit hôtel des monnoies de Lyon. Nous chargeons en conséquence, le commissaire du dit hotel des Monnoies de se concerter avec vous pour cet objet.

Les Commissaires Généraux des Monnoies
(signed) MAGINEL(?), MONGE, BERTHOLLET, LAGRANGE à MM. les Admin . . . Budgets de Rhône et Loire

[Document written by a scrivener in a conventional documentary script]

(*D*) Photostat of an unknown original entirely written by Lagrange

Me trouvant maintenant en vertu de l'arrêté du 19 thermidor an IV dans le cas de ne pas toucher l'indemnité accordée aux membres de l'Institut, je prie les Commissions des fonds de ne plus me comprendre dans les états de distribution de ces indemnités jusqu'à ce que je sois de nouveau dans le cas de les toucher.

À Paris ce 6 pluviose
an V(?) LAGRANGE

[The arrêté referred to in the letter was dated August 6, 1796, hence the letter was probably written on Jan. 5, 1797.]

(*E*) [February 22, 1799]

Le jury des écoles centrales du département de la Seine propose à l'administration centrale du département le citoyen Bouillon-La Grange, attaché à l'école polythecnique pour remplir la place de professeur de phisique, vacante par la mort du citoyen Saussure

Il propose également à l'administration centrale du departement, le citoyen Vincent peintre, membre de l'institut national, pour remplir la place de professeur de Dessin, vacante par la mort du citoyen Boullé.

A Paris, ce 4 ventôse an sept
(signed) LAPLACE, LAGRANGE, M. J. CHÉNIER
Séance du 6 ventôse
Adopté pour l'administration
Faire l'arrêté conforme

[The main part of this document above the signatures seems to have been written by Laplace. The three lines under the signatures are an administrative postil.]

(*F*) Undated. Possibly 1799 because of its likeness to (*E*)

Les membres du jury pour les écoles centrales de Paris, proposent à l'administration du département comme professeurs et bibliotécaires,
pour les mathématiques, le citoyen Costas
pour la phisique, le citoyen Libes
pour les langues anciennes, le citoyen Villars, membre de l'Institut
pour l'histoire, le citoyen Maherault
pour les deux biblioteques des écoles centrales qui n'en ont point encore, les citoyens Nicoleau et Eymar
(signed) LAPLACE, LAGRANGE
approuvé en séance du 11 th.dor [= Thermidor]
D. Pour (?) Eymar, Place St. Michel no. 785
Pour les Jésuites

Le citoyen St. Ange, Professeur de Belles Lettres Rue St. Apolline no. 6 destiné pour l'abbaye St. Martin des Champs (Cette maison ne devant pas servir à l'Ecole centrale, le citoyen St. Ange demande a être destiné pour la première Ecole qui ouvrira).

[Document written by various hands; the main part over the signature seems to have been written by Laplace.]

These six documents are not of great interest, but together they give us glimpses of various phases of Lagrange's life, (*A*) his social life, (*B* and *C*, 1792) administration of the Mint, (*D*) a dignified appeal concerning his unpaid salary, (*E* and *F*, 1799) educational administration.

We publish facsimiles of (*A*) and (*D*); a facsimile of (*F*) will be found in D. E. Smith, *History of Mathematics* I: 484, Ginn, 1923; *Isis* **6**: 440–444.

XXIII. BIOGRAPHICAL SOURCES

The main biographical source is Lagrange's correspondence, especially the letters exchanged

[110] This dating from 1789 is anterior to the Republican calendar which began in September 22, 1792. The document was written by a clerk.

with Fagnano, Euler, d'Alembert, Condorcet, and Laplace. Next come the accounts of people who knew him personally or were his contemporaries, to wit, Biot, 1813; Cossali, 1813; Delambre, 1816; Denina, 1790; Lacroix, 1815; Frédéric Maurice, 1814, 1819; Mirabeau, 1789; Thiébault, 1804; Virey and Potel, 1813. There are probably other companions of Lagrange in Turin, Berlin, or Paris, who have published remembrances of him; I shall be grateful to readers calling my attention to them. At any rate the letters and the accounts of those ten men give us a sufficient basis to recreate the essential traits of his personality.

A few books (*Annali*, Cantor, Nielsen, Pasquier) are listed for mathematical rather than biographical reference. Of course, every historian of mathematics, astronomy, mechanics, or physics dealing with Lagrange's time is obliged to discuss his works, but an enumeration of all those histories would be irrelevant.

Annali di Matematica. 1913. Ser. 3, XX, Aprile; XXI Ottobre. Milano.

These two volumes were published to celebrate the centenary of Lagrange's death. They contain only one historical paper (Loria's), but a few others discuss Lagrangian problems.

Biot, Jean Baptiste (1774–1862). 1813. Notice historique sur M. Lagrange. *Journal de l'Empire*, 28 avril. Reprinted in Biot's *Mélanges scientifiques et littéraires* III: 117–124, Paris, 1858.

Cantor, Moritz (1829–1920). 1908. *Vorlesungen über Geschichte der Mathematik* IV. 1759–1799. Leipzig, 1908.

Cossali, Pietro (1748–1815). 1813. *Elogio di L. Lagrange*. Padova.

Written promptly after Lagrange's death, for the Institut received a copy on August 30, 1813. Not seen by me but used by Delambre and Loria. Cossali was the author of a history of algebra, *Origine, trasporto in Italia, primi progressi in essa dell'algebra*. (2 v. Parma, 1797–1799).

Delambre, Jean Baptiste Joseph (1749–1822). 1816. Notice sur Lagrange. *Mémoires de la classe des sciences mathématiques de l'Institut*, II[e] partie, Paris. Reprinted in *Oeuvres* I: ix–li.

Denina, Carlo (1731–1813). 1790–1791. *La Prusse littéraire sous Fréderic II*. 3 v. Berlin. See II: 140–147, 1790.

Genocchi, Angelo. 1883. Luigi Lagrange. *Primo secolo della R. Acc. Torino*, 86–95. Torino.

Guareschi, Icilio (1847–1918). 1913. Notizie storiche intorno a Luigi Lagrange. *Mem. R. Acc. Torino* **64**, (1).

Lacroix, Sylvestre François (1765–1843). 1815–1855. Liste des ouvrages de M. Lagrange. Appended to *Mécanique analytique*, 2nd ed. II: 372–378, Paris, 1815. Reprinted in the 3rd ed. II: 383–389, Paris, 1855.

(Lagrange, J. L.). 1815. *Catalogue des livres de la Bibliothèque du Comte Lagrange*. Paris.

Not seen. Received by the Institut on March 13, 1815. (*Inst. de France, Procès.-verbaux* **5**: 466, 1815.)

Loria, Gino. 1913. G. L. Lagrange nella vita e nelle opere. *Annali di matematica* XX: ix–lii. Reprinted in Loria's *Scritti, conferenze, discorsi*, Padova, 1937; *Isis* **27**: 522–524.

——. 1938. Nel secondo centenario della nascita di G. L. Lagrange, 1736–1936. *Isis* **28**: 366–375, 9 ill.

Maurice, Frédéric (1775–1851). 1814. Directions pour l'étude approfondie des mathématiques recueillies des entretiens de Lagrange. *Le Moniteur universel*, Paris, 26 février.

——. 1819. Lagrange. *Biographie universelle*, XXIII: 157–175, Paris.

Mirabeau, Honoré Gabriel Riquetti, comte de (1749–1791). 1789. *Histoire secrette de la cour de Berlin, ou Correspondance d'un voyageur françois depuis le cinq Juillet 1776 [sic for 1786] jusqu'au dix-neuf Janvier 1787. Ouvrage posthume*. 2 v. Paris.

The book appeared anonymously without indication of place, publisher or printer; for additional security it was stated in the Avis de l'éditeur that the author "died last year in an unknown village in the depths of Germany." The author was Mirabeau, who lived until 1791.

——. 1895. *The secret history of the court of Berlin*. 2 v. London.

Indifferent translation of the previous item.

[Moscow]. 1937. Academy of Sciences. Collection of articles to celebrate the second centenary of Lagrange's birth, 1937. In Russian; contents in *Isis* **28**: 199, 1938.

Nielsen, Niels (1865–1931). 1929. *Géomètres français sous la Révolution*, 136–152. Copenhague; *Isis* **16**: 502; **30**: 514.

Pasquier, L. Gustave du. 1927. *Léonard Euler et ses amis*. Paris. *Isis* **11**: 456.

Riccardi, Pietro (1828–1898). 1873–1876. *Biblioteca matematica italiana dalla origine della stampa ai primi anni del secolo XIX*. II (1), col. 1–11. Modena.

Thiébault, Dieudonné (1733–1807). 1804. *Mes souvenirs de vingt ans de séjour à Berlin*. 5 v. Paris.

——. 1806. *Original anecdotes of Frederick the Great*. 2 v. Phila., Bronson.

Free and abbreviated translation of the preceding item.

Virey, Julien Joseph, et Potel. 1813. *Précis historique sur la vie et la mort de Lagrange*. Paris.

* * *

Bigourdan, Guillaume (1851–1932). 1901. *Le système métrique des poids et mesures*. Paris.

——. 1928. *Le Bureau des longitudes*. Notices appended to the *Annuaire du Bureau des longitudes*, Paris, 1928 and following years.

The early history of the Bureau will be found in the volumes for 1928, notice *A*, and for 1929, notice *C*.

Gauja, Pierre. 1934. *L'Académie des sciences de l'Institut de France*. Paris. *Isis* **25**: 304.

Harnack, Adolf von (1851–1930). 1900. *Geschichte der Königlich Preussischen Akademie der Wissenschaften*. 3 v. Berlin.

Maury, L. F. Alfred (1817–1892). 1864. *L'ancienne Académie des sciences*. 2[me] éd. Paris.

[Paris]. *Ecole Normale*. 1895. *Le centenaire de l'Ecole Normale*. Paris.

[Paris]. *Ecole Polytechnique.* 1894–1897. *Livre du centenaire, 1794–1894.* 3 v. Paris.

[Paris.] Institut de France. Académie des Sciences. 1910–1914. *Procès-verbaux des séances de l'Académie tenues depuis la fondation de l'Institut* I–V [1795 to 1815]. Hendaye, Basses-Pyrénées.

[Turin]. 1883. *Primo secolo della R. Acc. delle scienze di Torino, 1783–1883.* Torino.

XXIV. ICONOGRAPHY

Lagrange always refused with some impatience to sit for a portrait, being convinced (says Virey, p. 12) that only the qualities of the spirit or the soul deserve to be perpetuated in the memory of men. However, a sketch was made while he was dozing, probably at a meeting of the Institut, and after his death a cast was taken of his face.

That "heureux larcin" to use Virey's phrase must have been repeated many times, for we have various portraits representing him at different ages. Four of them, plus a sculptured medallion, may be easily compared in Loria's article (*Isis* **28**: 366–375, 1938).

All are restricted to the head and bust, or head and shoulders. The best known is one showing him in old age (Loria, fig. 6); this is probably the one referred to by Virey and it is the one which I am inclined to believe is the most faithful. It is an engraving by Achille Martinet (Paris, 1806–Paris, 1877) from a painting or drawing by François Joseph Heim (Belfort, 1787–Paris, 1865). We reproduce it in frontispiece.

Loria's article contains also a photograph of the very effective but conventional statue which was erected in one of the public squares (close to the Piazza Carlo Felice) of Lagrange's native city.

According to Maurice (1819: 174 footnote) there was a bust of Lagrange in the Chambre des pairs, very like him though made after his death.

The illustrations of this article, other than Lagrange's portrait, require no explanation. They are all due to the courtesy of the Harvard Library, except figure 11 which we owe to the kindness of Dr. Rodolfo Garcia, Director General of the National Library, Rio de Janeiro, figure 14 to the Pierpont Morgan Library, courtesy of its Director, Miss Belle da Costa Greene, figure 15 to the David Eugene Smith Library, Columbia University, courtesy of its keeper, Miss Bertha Frick.

THE LIFE PORTRAITS OF JEFFERSON AND THEIR REPLICAS

FISKE KIMBALL

Director, Philadelphia Museum of Art

(*Read April 21, 1944*)

THE National Gallery of Art, in its Jefferson Bicentennial Exhibition,[1] did a great service both to art and to history in assembling in one place so many of his principal portraits, including especially those painted from life. For the first time it really became possible to confront the conceptions of various artists, to compare their several versions, to know how Jefferson looked at various periods, and to determine which were the portraits painted from life.

Like Washington, Jefferson had the good fortune to be portrayed by the greatest sculptor of his time, Houdon, a supreme master precisely in the field of portraiture, and by a painter, Stuart, who among Americans remains unsurpassed in this field. Other artists of merit, including notably Charles Willson Peale, John Trumbull, and Thomas Sully, also numbered Jefferson among their sitters.

It was Charles Henry Hart, a half century ago, who first put in order the portraits of Jefferson, along with those of other founders of the Republic. In 1887, he organized in Philadelphia[2] a great assemblage of historical portraiture. Another in 1889, in New York, on the occasion of the centennial of Washington's inauguration, was supplemented by a richly illustrated memorial volume in folio,[3] illustrating many additional portraits. Shortly afterwards, in *McClure's Magazine* for 1898,[4] Hart published a paper on the life portraits of Jefferson, discussing and reproducing a number of them. These two publications have been the sources of most subsequent statements regarding the Jefferson portraits,[5] which have received but small attention compared with those of Washington.

It is time now to review the subject afresh, as the bicentennial occasion has permitted. Hart's careful search of the documents by no means exhausted them, nor did it prevent him from falling into a number of errors. Jefferson's pocket account books, to which he had access when they were in possession of Jefferson's great-granddaughter, Sarah Nicholas Randolph, have now largely come into public hands[6] and can be examined more at leisure. Jefferson's correspondence has also been more fully published and made accessible. Meanwhile there has been much further study of the individual artists. Houdon, Stuart, Sully, and Trumbull have all been made the object of important monographs, as has Charles Willson Peale, though this has not yet reached the period when he painted Jefferson. We shall find that, even beyond these, there is still new light to be thrown on the portraiture of Jefferson.

We shall pass over the several "profiles" or silhouettes of Jefferson of which we have record,[7]

---

[1] National Gallery of Art, *The Thomas Jefferson Bicentennial Exhibition*, 1943.

[2] Pennsylvania Academy of the Fine Arts, *Loan Exhibition of Historical Portraits*, 1887. Hart was chairman of the Committee on Exhibitions and signed the Introductory Note.

[3] *History of the Centennial Celebration of the Inauguration of George Washington*, ed. by C. W. Bowen, Appleton, 1892. Chapter xxii is "Notes on Portraits, by the Editor." He states in the Preface: "To Charles Henry Hart of the Art Committee . . . the Publication Committee is also indebted. Much of his time during the past two years has been spent in giving such information about portraits as it was impossible to obtain elsewhere."

[4] Life Portraits of Thomas Jefferson, *McClure's Magazine* **11**: 47–55, 1898.

[5] This is largely true even of the section on "Portraiture of Thomas Jefferson" by John Hill Morgan in the volume *Virginia Historical Portraiture*, ed. by A. W. Weddell, Richmond, Byrd, 1930.

[6] 1767–1770, Library of Congress (lent by General Jefferson Randolph Kean); 1771–1772, Huntington Library; 1773, Library of Congress; 1774, Massachusetts Historical Society; 1775, Huntington Library; 1776–1778, Massachusetts Historical Society; 1779–1782, Library of Congress; 1783–1790, Massachusetts Historical Society; 1791–1803, New York Public Library; 1804–1826, Massachusetts Historical Society.

[7] We find Jefferson in Paris paying on April 23, 1789, "for taking profile, 6 livres," and again, six days later, "pd. for profile, 30 livres." Neither of these profiles is known to exist, though we may confidently expect some, true or false, to appear soon. Jefferson's account book for 1804 contains the following entries regarding silhouettes,

and the paintings reputed to stand back of various engravings,[8] of which we know nothing further. We shall devote our attention entirely to the portraits by painters and sculptors known from documents to have portrayed Jefferson from life.

This, as we shall see, will limit us to portraits by the following artists, regarding whom we either find mention of payments in Jefferson's accounts or have other documentary evidence that Jefferson actually sat to them: Mather Brown, 1786; Trumbull, 1786 and 1787; Houdon, 1789; Giuseppe Ceracchi, 1790 or 1791; Charles Willson Peale, 1791; William Joseph Williams, 1792; Gilbert Stuart, 1800 and 1805; Saint-Mémin, 1804; Rembrandt Peale, 1805; Bass Otis, 1816; William J. Coffee, 1818; Sully, 1821; and, for a life mask, John Henri Isaac Browere, 1825—thirteen painters and sculptors in all.

This leaves aside a number of well-known representations of Jefferson made during his lifetime, such as, among others, the oil by Charles Peale Polk, the wax by Valaperta, the water-color by Robert Field, the plaster relief by George Miller, and the bust by Pietro Cardelli,[9] as well as, indeed, the Sharples pastels—since we do not know definitely that any of these men had life sittings. It would not be impossible that, as in the case of Field's miniatures of Washington, some of them were derived and modified from portraits by others.

On the other hand, it would be improper not to add the pencil sketch by the architect Benjamin Henry Latrobe who, though he scarcely had a formal sitting, appears to have sketched Jefferson from life, as their intimate collaboration over a considerable period of years readily permitted. This makes fourteen artists all told who surely portrayed Jefferson from life.

It may seem somewhat strange that Jefferson, who had achieved celebrity at thirty-one, as author of the *Summary View of the Rights of British America*, and immortality at thirty-three, as author of the Declaration of Independence, should not have sat to anyone for his portrait before he reached the age of forty-three. This was certainly not because of lack of predilection for art.[10] Edmund Randolph said that Jefferson from his youth "panted after the fine arts."[11] It was far from being the result of any lack of financial means or social standing. Contrary to the loose expressions of some careless writers, and to the historical construction fostered by writers of the "frontier school," Jefferson was by no means a backwoodsman of pioneer stock. Not merely through his mother, Jane Randolph, but through his father, Peter Jefferson, he came from the class of the landed Tidewater planters, well established in Virginia for more than a hundred years before his birth. Randolphs and Jeffersons alike, like all the great landed families, moved gradually up the rivers as the wasteful culture of tobacco exhausted their older fields. It was only six years before Jefferson's birth that his parents moved from the plantation of Fine Creek—opposite Tuckahoe on the James, a little above Richmond—which Peter Jefferson had inherited from his father, to Shadwell on the Rivanna in Albemarle County, itself far inside the current frontier of the Alleghenies. Two years after he was born, they returned to the Tidewater, and of Jefferson's life from then to the age of twenty-four only three years together were spent in the Piedmont. The true status of the young Jefferson is that of an aristocratic youth, admitted to the familiar society of Governor Fauquier, and on terms of the closest intimacy with the Pages, Burwells, Lewises, Willises, Harrisons, and other scions of the great Tidewater plantations.

Jefferson's youth, however, was the moment when Virginia was most poorly provided with portrait painters. John Wollaston, who had painted older Randolphs from 1755, when Jefferson was a boy of twelve, left Virginia for good in 1767,[12] just as Jefferson was admitted to the bar. It is interesting to observe that the portraits of Jefferson's friends and cronies, Warner Lewis, Thomas Nelson, Jr., and John Page,[13] were

which have been assumed to be of himself: "Jan. 7, gave Mr. Doolittle order on J. Barnes for 10 D. for 2 profiles," and "Feb. 10, gave Doolittle ord. on J. Barnes for 10 D. for profiles."

[8] Some of these are listed and discussed in the *History of the Centennial Celebration of the Inauguration of George Washington*, especially on page 489.

[9] Even its identity still remains unknown, as in Hart's time. Conceivably it may be of the type of a bust found at Edgehill after 1902 (*cf. Annual Report of the Monticello Association*, 13, 1938), of which numerous casts were subsequently made, one being since 1942 in the Lane High School in Charlottesville.

[10] *Cf.* the writer's "Jefferson and the Arts," *Proc. Amer. Philos. Soc.* **87**: 238 ff., 1943.

[11] Essay on the Revolutionary History of Virginia, *Virginia Magazine*, **42**: 122, 1938.

[12] T. Bolton and H. L. Binsse. Wollaston, An Early American Portrait Manufacturer, *Antiquarian:* 30–34, 50–52, June 1931.

[13] Illustrated in *Virginia Historical Portraiture*, cited above.

all done in childhood, when Jefferson, as a minor and ward, had not yet come into his fortune, and before they all went to college together. In the time around 1770, the only painter known to have worked in Virginia was John Durand, of whose works—"hard and dry," as William Dunlap wrote—the portraits of Thomas and Amy Newton are dated in that year.[14]

Charles Willson Peale, from his return from London in 1769, was seated in Maryland with annual trips to Philadelphia; his only early sitter in Virginia being Washington in 1772, at Mount Vernon, merely across the Potomac. Matthew Pratt was in Williamsburg a fortnight in March of 1773, at a moment when Jefferson was briefly there; indeed, Pratt then painted at Tuckahoe the infant son of Jefferson's cousin. Perhaps Jefferson's hopes already ranged higher. In the summers of 1774 and 1775, Peale painted in Williamsburg. During the whole summer of 1774, however, Jefferson was in the Piedmont. In 1775, he was in Congress at Philadelphia, and reached Williamsburg only for ten days in mid-August. Perhaps his purchase, in those very days, of John Randolph's fine Cremona violin, was a sufficient extravagance for one moment.

In 1766 Jefferson had made a brief journey to the northward, primarily to be inoculated in Philadelphia against the smallpox. He visited Annapolis, on his way, and pressed on for a few days in New York. It was then a half dozen years since Benjamin West had gone from Philadelphia to Rome and London. Matthew Pratt had left for London in 1764 and did not return until 1768. Copley was not to visit Philadelphia and New York until 1770; he left America for Europe in 1774.

It has scarcely been realized also how briefly Jefferson was in the North, even at the time of the Revolution. He was not a member of the first Continental Congress of 1774. In the second he took his seat on June 20, 1775, departing on its adjournment August 1. He arrived for the fall session in October. On December 28, alarmed over his wife's health, he left for Virginia, returning on May 14 and remaining only until September 3.[15] From that autumn he sat in the House of Delegates in Virginia, at work on the famous revisal of the Virginia laws. In 1779 to 1781, in the darkest days of the Revolution, he was governor of the State.

In 1782 came his wife's death. Then only was he briefly again in the North, in Philadelphia from December 26 to January 26, 1783, and again October 29 to November 21, then in Congress at Annapolis until May 11, 1784. After a fortnight in Philadelphia, he made a rapid trip through New England before sailing from Boston on July 5, 1784, on a mission to Europe.

One cannot say, of course, that there were no opportunities, in this time, for Jefferson to have been painted. It was just before his brief stay in Philadelphia in the winter of 1782–1783 that Peale had completed his new exhibition room, "ornamented with the portraits of a great number of worthy personages," including Washington and the French officers. Yet it was only at the end of 1791 that Jefferson was to be added to this gallery. When passing through Philadelphia on his way north in May, 1784, Jefferson turned, not to Peale, but to Joseph Wright, lately arrived from England, for a portrait of Washington to take with him to Paris, as material for the statue of Washington which was being considered by the Virginia Assembly.[16] This work, a replica of the one painted by Wright from life the year before, had itself to be sent off unfinished, so brief was the time available.

Thus it was not until nearly two years after Jefferson's arrival in Europe that, for the first of many times, he sat for his own portrait. Although he early formed connections in Paris with artistic circles, and notably with Houdon, his earliest portrait was not made in Paris, nor by a European artist, but in London, and by an American—an American less celebrated than West—the young Mather Brown, then in his first fashionable success.

Jefferson went from Paris to London in March 1786, arriving the 11th and leaving on April 26. The day before his departure, as we find in his

[14] *William and Mary Quarterly* **2**: 31, 1892.

[15] Charles Willson Peale was in Annapolis through most of 1775; he arrived in Philadelphia December 19, staying through January; he returned March 1, staying into April, and again in May, when on the 18th he rented a house and was commissioned next day by Hancock to paint the Washingtons. Adams visited his painting room August 11, and met Francis Hopkinson there, but we hear nothing of any visit from Jefferson, either in Peale's diary or Jefferson's accounts. C. C. Sellers, *The Artist of the Revolution: The Early Life of Charles Willson Peale*, 127–134, Hebron, Conn., Feather & Good, 1939.

[16] Fiske Kimball, Joseph Wright and His Portraits of Washington, *Antiques* **15**: 377–382, 1929. The state of Virginia commissioned a full length portrait of Washington by Peale for the purpose, but this was not completed until October 30 and reached Jefferson only on April 30, 1785, after Houdon had already undertaken to come to America.

FIG. 1. Jefferson, by Mather Brown, 1786. Replica, printed in London for John Adams, of the life portrait painted for Jefferson, 1786. Charles Francis Adams, Esq., Quincy, Massachusetts. Photograph by courtesy of the National Gallery of Art.

accounts, he "pd, Brown for my picture £10." This picture we know only through the replica (fig. 1)[17] a seated half-length, preserved in the Adams family, for which the receipt, pasted on the back of the stretcher, reads: "London, May 12, 1786. Received of his Excellency, John Adams, Esq., six guineas for a kit-cat portrait of Mr. Jefferson. Mather Brown." An item in the *New-York Packet*, October 19, 1786, speaks of this picture as visible in Brown's room in London, "among pictures of near one hundred Americans."[18]

This earliest Jefferson portrait, little known to the general public, gives a favorable idea both of the ability of the artist and of the attractiveness of the subject. It is freely painted with fresh color in the ruddy countenance and in the ruffles of striped reddish-brown and white. While not conventionally handsome, Jefferson shows the engaging vivacity and intelligence which made him a favorite in the Paris salons. The hair, his own,[19] is powdered, with rolls over the ears. In the background is a classical figure with a staff crowned by a Phrygian cap which has correctly been interpreted as a statue of liberty.

There is an allusion by John Trumbull to this Jefferson portrait, as well as to a portrait of Adams which Brown was painting for Jefferson, as to be forwarded soon, in an unpublished letter from London to Jefferson in Paris, March 6, 1788: "Mr. Adams is like. Yours I do not think so well of." Jefferson replied May 18: "You say mine does not resemble. Is it a copy? Because he agreed that the original should be mine, and it was that I paid him for." Trumbull answered May 23: "I believe what he means to send you of yourself to be the copy and that of Mr. Adams thus the original. They will now be sent you. . . ." Nevertheless, in view of the smaller payment by Adams we must suppose Adams did not receive the original Jefferson. Presumably Jefferson duly received these pictures, for his portrait of Adams, as well as a portrait of Tom Paine which he had commissioned Brown to paint, were exhibited by the Boston Athenaeum in 1828, with the bulk of Jefferson's collection.[20]

What became of Jefferson's original life portrait by Brown we can only surmise. There is no reference to it in any of the descriptions of Monticello, which mention many other works of art. Neither is there any allusion to it in the lists and correspondence regarding the dispersal of Jefferson's collection after his death.[21] As to the possible fate of this and other missing possessions, we have only two clues. When Jefferson retired from the Presidency, his things were sent round by water. From a letter regarding the loss of his collection of Indian vocabularies,[22] we know that at least one trunk was broken open by thieves, and its contents thrown into the James. Conceivably this may have happened to other boxes. Another possibility is that Jefferson may have given the portrait to his granddaughter Ellen Wayles Randolph, who married Joseph Coolidge of Boston in 1825. Jefferson wrote her November 14, 1825: "We have heard of the loss of your baggage, with the vessel carrying it;" and March 20, 1827, she wrote her sister, Mrs. Trist: "All that I had myself received except the desk and watch and a few letters are buried in the ocean." Possibly the picture may still exist in some branch of the Jefferson descendants, though their possessions have been pretty fully canvassed. If not, I presume we may soon be subjected to the appearance of such a picture, based on the Adams replica, and complete with supposed pedigree. It had better be a good one.

It was while in London that Jefferson became acquainted with Trumbull, then thirty years of age, who had been studying with West since 1780, and was already meditating his project for a series of paintings on "subjects of national history." Trumbull spent the summer and early fall of 1786 in Paris, seeing much of Jefferson. "In the autumn of 1787," he wrote, "I again visited Paris, where I painted the portrait of Mr. Jefferson in the original small Declaration of Independence."[23] In a letter to Jefferson December 26, 1816, he wrote: "Twenty-eight years have elapsed since

[17] Oil on canvas, 35¾ × 28⅛ in., relined. Signed and dated, lower right, "M. Brown Pt. 1786." Property of Mr. Charles Francis Adams.

[18] The notice is published in William Kirby's *Notes on American Artists*, N.-Y. Hist. Soc., 29, 1922.

[19] There are no payments for wigs in Jefferson's accounts.

[20] The catalogue of the exhibition, published by M. M. Swan, *The Athenaeum Gallery*, 86–87, Boston Athenaeum, 1940, erroneously lists the Paine as by Trumbull.

[21] It does not appear in Jefferson's manuscript catalogue of paintings—preserved by his descendant Mrs. Edwin Kirk, who has kindly permitted it to be transcribed—but that is manifestly incomplete, and does not include any of the portraits of Jefferson himself. The collection was first shown in New York, and the portrait may possibly have been sold there, though again we judge that portraits of Jefferson were not included in the works publicly offered.

[22] *The Writings of Thomas Jefferson*, ed. by Albert Ellery Bergh and Andrew A. Lipscomb, XIV: 402, Washington, Thomas Jefferson Mem. Assn., 1907.

[23] John Trumbull, *Autobiography, Reminiscences and Letters*, 150, Wiley & Putnam, 1841.

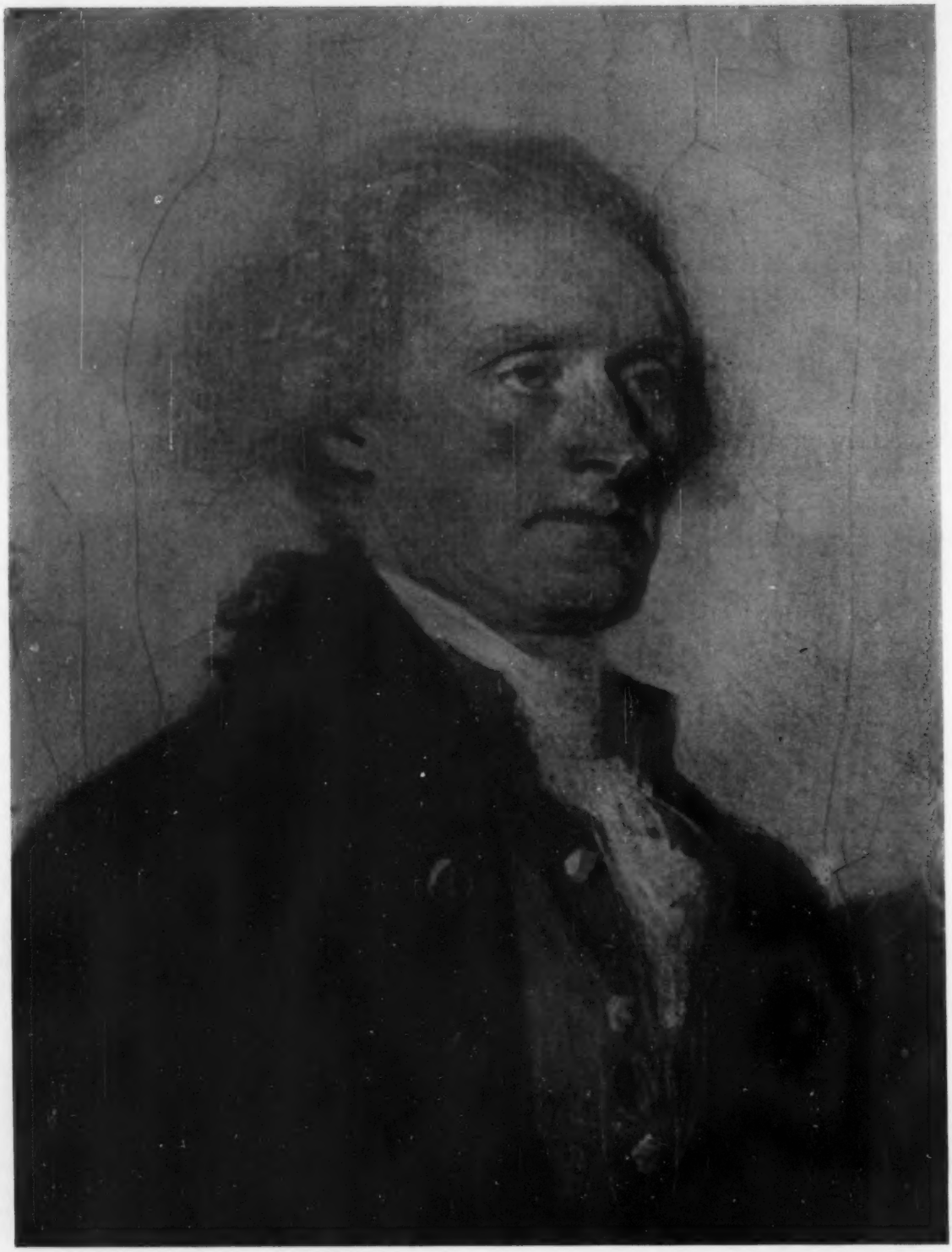


FIG. 2. Jefferson, by John Trumbull, painted from life 1787. Detail from the Declaration of Independence. Yale Gallery of Fine Arts.

under the kind protection of your hospitable roof at Chaillot, I painted your portrait in my picture of the Declaration of Independence, the composition of which had been planned two years before in your library."[24] This would make the date 1788, which is incorrect, as Trumbull was not in Paris after 1787. It is important to stress these statements, for they show that the Jefferson in the painting at the Yale Gallery of Fine Arts (fig. 3) is a life portrait. It has hitherto scarcely received the attention it deserves, as will be appreciated by the detail here shown (fig. 2).

The carriage of the head is of great nobility, fearless and proud, as well it might be in the moment of this pledge of "our lives, our fortunes, and our sacred honour." Resoluteness and courage are in the firm chin and set mouth; intelligence shines from the clear eyes.

In this picture Jefferson's fine hair has its natural reddish color. We know that in 1786 he wore it powdered, as in the Mather Brown portrait, but Trumbull was eager to show the actors in the scene as they had been—he speaks of having taken the opportunity of painting Adams in 1787 when he had combed the powder out of his hair, before returning to America. Jefferson wears a red waistcoast, as he may well actually have done on the occasion—five were in his wardrobe on his marriage in 1772.[25]

The Trumbull painting is also of interest as one of but two by artists who painted him from life showing Jefferson's figure at full length—the other being that of Sully, over thirty years later. According to his grandson, Thomas Jefferson Randolph, so intimate with him, "Jefferson's stature was commanding, six feet two and a half inches in height, well formed, indicating strength, activity and robust health; his carriage erect."[26] So he appears in Trumbull's painting, towering above the other members of the committee.

There descended in the family of Jefferson an oil miniature of Jefferson[27] by Trumbull (fig. 4). Thomas Jefferson Trist, a great-grandson, bequeathed this to his sister, Mrs. John W. Burke of Alexandria, Virginia, who died in 1915.[28] When in her possession, in 1892, it was reproduced and discussed by Bowen,[29] who supposed it to be the original life portrait, and remarked that "the head of Jefferson in the 'Declaration of Independence' is not a close copy of the original." Actually, we believe, the miniature is a modified replica made by the artist to leave with Jefferson, his host. It is more hasty and sketchy in execution, following the Declaration portrait in the pose of the head and in showing the unpowdered hair, but with the body frontal and differences in the costume.

Another oil miniature of Jefferson (fig. 5)[30] was bequeathed to the Metropolitan Museum of Art in 1924 by Cornelia Cruger, and was at first listed as "attributed to John Trumbull,"[31] later as by Trumbull.[32] She was a descendant of Catherine Church (Mrs. Bertram P. Cruger) who, like her mother, Angelica Schuyler Church, was a close friend and correspondent of Jefferson. Jefferson wrote to Angelica Church on August 17, 1788:

> The memorial of me which you have of Trumbul [*sic*] is the most worthless part of me. Could he paint my friendship to you, it would be something out of the common line.[33]

In a letter of 1820 to Maria Cosway,[34] he speaks of Mrs. Cruger in the same breath with Trumbull, as members of "our former coterie" in Paris. The authentic provenance of the picture is thus well established, it is from Trumbull's own hand.

---

[24] MS at Yale University Library, published by kind permission of the Library Committee and of Mr. Theodore Sizer.

[25] "List of our clothes," made by Mrs. Jefferson, published by Marie Kimball, *Jefferson: The Road to Glory*, 173, N. Y., Coward-McCann, 1943.

[26] Letter to H. S. Randall, published in Randall's *The Life of Thomas Jefferson* III: 675, N. Y., Derby, 1858.

[27] Oil on panel, oval, $4\frac{7}{8} \times 3\frac{3}{4}$ in.

[28] It passed then to her nephew, Edmund Jefferson Burke, and now belongs to his estate. At some time since 1892 the face of the picture was lamentably scratched by a child.

[29] *History of the Centennial Celebration of the Inauguration of George Washington*, facing p. 21, and text p. 486.

[30] On panel, $4\frac{7}{16} \times 3\frac{1}{4}$ in., enclosed in an oval within the rectangle. On the back of the frame is inscribed "Thomas Jefferson / owned by John C. Cruger," i. e. John Church Cruger, the father of Cornelia Cruger. Mrs. H. D. Allen of the Metropolitan Museum staff informs me that Miss Cruger left by will "a small book *Télémaque* given by Thomas Jefferson to Catherine Church."

[31] *Bulletin of the Metropolitan Museum of Art* **19**: 53, 1924.

[32] H. B. Wehle, *American Miniatures*, 28, Garden City, 1927, where it is reproduced on plate 11; Metropolitan Museum exhibition catalogue *Franklin and His Circle*, No. 99, 1936. Mrs. H. D. Allen of the Metropolitan Museum has kindly supplied further information regarding the picture.

[33] *Thomas Jefferson Correspondence printed from originals in the collections of W. K. Bixby*, 32–33, Boston, 1916.

[34] December 27, 1820, *The Writings of Thomas Jefferson*, ed. by A. E. Bergh and A. A. Lipscomb, XVIII: 309.

FIG. 3. The Declaration of Independence. Detail from the painting by John Trumbull. Yale Gallery of Fine Arts.

FIG. 4. Jefferson, oil miniature by John Trumbull, 1787. Estate of Edmund Jefferson Burke, Chestnut Hill, Massachusetts. Photograph by courtesy of the National Gallery of Art.

FIG. 5. Jefferson, oil miniature by John Trumbull, 1786. Metropolitan Museum of Art.

One striking difference between this Church miniature and the other Trumbull heads of Jefferson is that, whereas in them (as well as in the Houdon bust of 1789) the hair is drawn loosely back over the ears, in it the hair, which is powdered, is in horizontal rolls above the ears, as in Mather Brown's portrait of 1786. We can scarcely resist the conclusion that the Church miniature was painted in 1786,[35] a year before the Declaration and Burke portraits. If so, it is the second in date among the life portraits of Jefferson.

Early in his stay in Paris, Jefferson had learned "that a Monsieur Houdon, of this place, possesses the reputation of being the first statuary in the world."[36] After engaging him to make for Virginia the statue of Washington, and a bust of Lafayette, also commissioned by the state, he wished, very naturally, to possess some of his works. Houdon had earlier made a bust of Paul Jones, and Jones in 1786 presented one of numerous plasters of this to Jefferson, though Jefferson had ultimately to pay the bill for it. At Monticello in Jefferson's later years[37] were, in addition, busts by Houdon of the following worthies: Washington, Franklin, Lafayette, Voltaire, and Turgot.

Hart's assumption that Houdon took a life mask of Jefferson rests on an allusion by Jefferson many years later. "Browere," he wrote, ". . . said his operation would be of about 20 minutes and less unpleasant than Houdon's method."[38] Jefferson may indeed have known Houdon's method from personal experience, but he could also have known it from observation. Houdon by no means always, or even often, employed it: only two or three indubitable instances are known.[39] We have no certain knowledge that he used it in the case of Jefferson.

In the Salon of 1789 was shown a plaster bust of "M. Sesseron, envoyé des États de la Virginie," and one remained in Houdon's studio at his death in 1828. Jefferson's accounts have the following entry on July 3, 1789: "Gave Houdon order on Mr. Grand for 1000 livres for busts made for me." Hart misquoted and thus misunderstood this passage. He wrote: "for a bust in marble Jefferson paid Houdon . . . one thousand francs. . . . It was destroyed, by the carelessness of workmen at Monticello during Jefferson's lifetime." The passage in the accounts is perfectly clear and legible; the word "busts" is in the plural. Now 1000 livres or francs was quite inadequate for one bust in marble, let alone anything further. For the block of marble alone for the Lafayette bust, Jefferson advanced on behalf of the state 50 louis d'or, or 1200 francs. No matter how friendly Houdon was with Jefferson, he would scarcely have executed his bust in marble for less than actual cost, and 1000 francs would not have covered the cost of material, to say nothing of stonecutters' wages. On the other hand, additional plasters from existing models were not more than 100 francs.[40] Jefferson, as we have just seen, got at least five of other subjects, and we shall see that he must have got several of his own bust. That would account for the 1000 livres "for busts made for me."[41]

Jefferson thus never had a Houdon bust of himself in marble, and we do not need the legend of its destruction by a workman. Today two original plasters of Houdon's Jefferson, both signed *Houdon F*, are known to survive. One (fig. 6) repaired after being somewhat broken

[35] This dating and others proposed in this paper were communicated to the authorities of the National Gallery of Art and adopted by them in the check-list of works in the exhibition. The authorities at Yale have kindly looked in the Trumbull correspondence there, both at the Library and the Gallery, to see if the date can be positively confirmed by any correspondence between Trumbull and Miss Church, but without success.

[36] Letter to Washington, December 10, 1784. C. H. Hart and E. Biddle, *Memoirs of The Life and Works of Jean Antoine Houdon, The Sculptor of Voltaire and of Washington*, 186, Phila., 1911.

[37] Preserved at Monticello is a plan (expressly stated to be incomplete) made by Cornelia Jefferson Randolph, Jefferson's granddaughter, to show the location of works of art. A reproduction without the full legend of the original appears in S. N. Randolph's *Domestic Life of Thomas Jefferson*, 334, Harper, 1871. These show the Voltaire and the Turgot. The Washington, Lafayette, Franklin and Paul Jones busts, as well as perhaps the Voltaire, passed to the Boston Athenaeum through Jefferson's grandson-in-law, Joseph Coolidge, and the first four still exist. *Cf.* M. M. Swan, *The Athenaeum Gallery*, 161–167, 1940.

[38] Letter to Madison, October 18, 1825, *The Writings of Thomas Jefferson*, ed. by A. E. Bergh and A. A. Lipscomb, XIX: 287, 1907.

[39] G. Giacometti, *La Vie et l'oeuvre de Houdon*, II, 2–3, 34, 116, 173–174, Paris, Camoin, 1929. C. H. Hart and E. Biddle, *op. cit.* 201.

[40] In 1775 Houdon contracted with Sophie Arnould for a marble and thirty plasters at a total of 3,800 livres, with the privilege of ordering additional plasters at 60 livres apiece. G. Giacometti, *op. cit.*, I, 133. Coffee in 1785 charged Franklin 96 francs for such additional busts. *Cf.* C. H. Hart and Edward Biddle, *op. cit.*, 88.

[41] The packers' lists of things sent from Paris to Philadelphia in 1790 lists only five busts in all, but conceivably some of himself had already gone to America with his baggage in 1789.

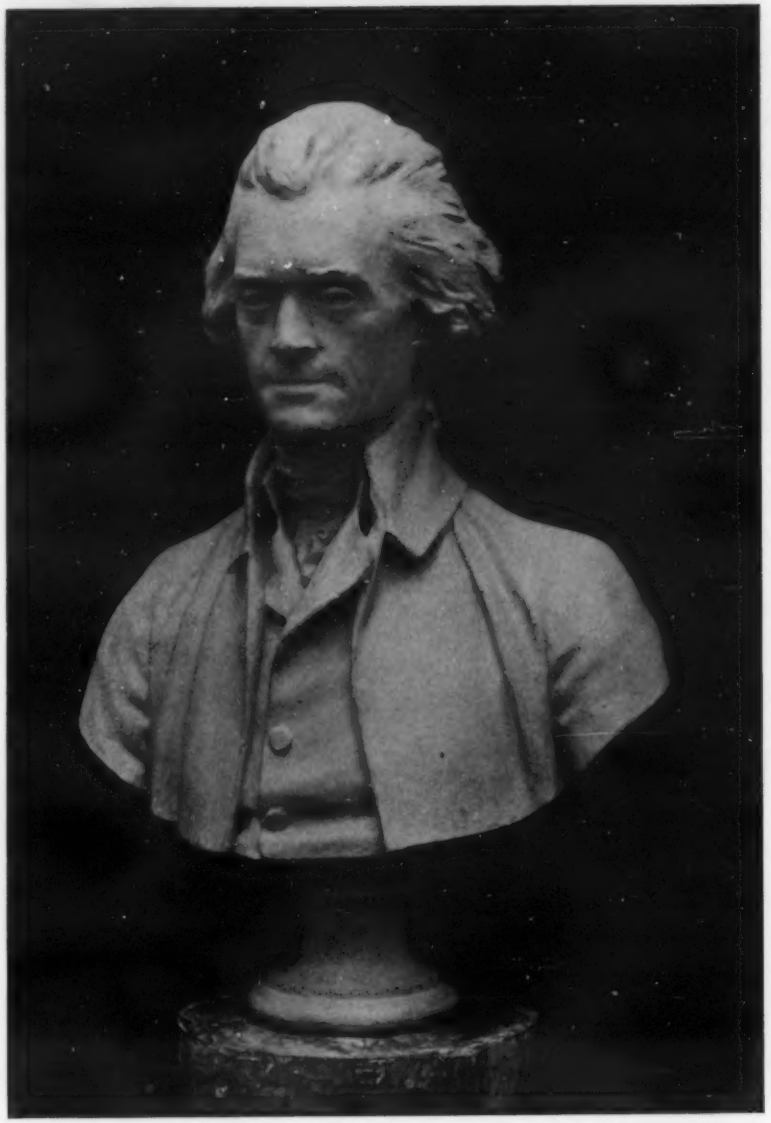


FIG.6. Jefferson, by Houdon, 1789, plaster. American Philosophical Society.

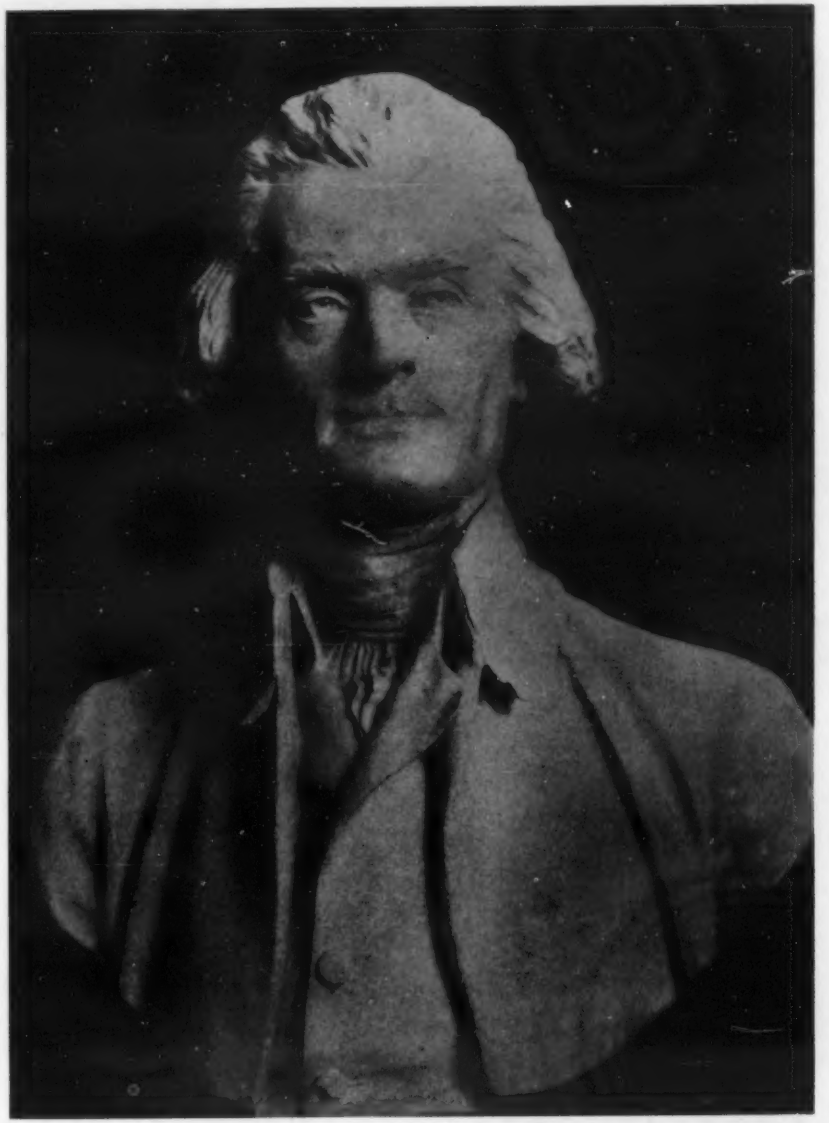


FIG. 7. Jefferson by Houdon, 1789, plaster. New-York Historical Society.

long ago, belongs to the American Philosophical Society, where it was already to be seen in 1819.[42] This bust is said to have been given by Jefferson to Rittenhouse. The other (fig. 7) unbroken, belongs to the New-York Historical Society, to which it was given by Mrs. Laura Wolcott Gibbs in 1839. This one is said to have been given by Jefferson to Hugh Williamson. Now while Rittenhouse and Williamson were friends and correspondents of Jefferson, they were by no means his closest friends. No mention of such gifts appears in his correspondence with them. On the other hand, no busts of Jefferson are known to have been in the possession of Washington or of Madison,[43] who would have been much more apt to receive them. Moreover the plan of Monticello prepared on the dictation of Jefferson's granddaughter Cornelia Jefferson Randolph shows but a single bust of Jefferson, in the hall, and that would have been the marble bust by Ceracchi. What became of any plaster Jefferson himself may have kept of his bust by Houdon remains a mystery. The same possibilities of loss exist as in the case of the Mather Brown portrait.

Houdon's bust of Jefferson, especially as it appears in the plaster at the New-York Historical Society, is a work of the highest quality, among the most realistic portraits of the artist. This is the Jefferson of the French Revolutionary era, with eye of fire, filled with enthusiasm for the rights of man.

About 1934 there appeared in the art market a marble Houdon Jefferson, which was purchased in that year by the Boston Museum of Fine Arts. It was reputed to have come from the family of Destutt de Tracy.[44] Writing to Jefferson in 1802, Lafayette speaks of Senator Tracy as one whom Jefferson had known as a patriot member of the Constituent Assembly.[45] Jefferson's correspondence with Destutt began in 1806. The suggestion was also made that the bust "was made originally for Lafayette. . . . Lafayette's daughter married into the Destutt de Tracy family, and it is believed that through inheritance from her father she may have brought the bust into the Destutt collection."[46] Possibly, but did Lafayette himself ever own a marble Jefferson? There seems to be no reference to one in all the many letters which passed between them, both before and after Jefferson left Paris. It would, of course, have been perfectly possible for either Lafayette or Destutt to have secured a marble from Houdon at any time, after the plaster which remained in Houdon's studio at his death.

I would not have been inclined to raise any question as to the history of this bust, were it not that, at just the same period of its appearance, several other Houdon marbles of American subjects, hitherto unknown—including two Washingtons and a Franklin—beside numerous bronzes and plasters reputed to be antique, also appeared in the art market, some of them with signatures or dates obviously erroneous, and some of them under circumstances very difficult to explain.

The Museum of Fine Arts, however, states: "We had the Jefferson bust put through the tests in our laboratory before it was purchased. We found in its favor evidence of chemical and physical changes due to its exposure to light, atmosphere and to care and minor abrasions." Clearly this laboratory evidence should carry much weight in favor of the age of the bust. The testimony of sculptors, however, is that the treatment of the planes is not characteristic of the marble cutters of Houdon's own studio.

Regarding the portrait by Charles Willson Peale, the following unpublished letters to Jefferson are supplied by the courtesy of his descendant and biographer, Charles Coleman Sellers:

Sir

I was under a mistake in expecting the favor of your setting at One O'clock this day. You will

---

[42] William Short to Jefferson, October 21, 1819, *Mass. Hist. Soc. Collections*, ser. 7, 1: 289, 1900.

[43] Madison's collection in 1832 is described in J. E. Semmes, *John H. B. Latrobe, and his Times* 240, Baltimore, Norman, Remington, 1917; an inventory of Mrs. Madison's belongings November 15, 1842, is preserved in her papers at the Library of Congress.

[44] Mr. Edwin J. Hipkiss, who published this bust in the *Bulletin of the Museum of Fine Arts* **32**: 69–74, 1934, has very kindly informed me that the bust was acquired through the Marie Sterner Gallery, and has summarized the papers received with it, as follows:

"Letter of expertising by J. L. Souffrice, 1932;
Letter of expertising by Giacometti, 1929;
Letter of P. Leclercq who sold it in 1928;
Civil paper of 1868 showing that Ferdinand Leclercq inherited the bust from
Louis Leclercq, who married in 1839
Comtesse Sarah de Destutt, daughter-in-law of M. Destutt, Chateau de Paray, L'Orne, the friend of Jefferson."

[45] G. Chinard, *Jefferson et les idéologues*, 215, Johns Hopkins, 1925. Kindly called to my attention by Mr. Chinard.

[46] In the notice of the acquisition in the *New York Times*, September 30, 1934, citing Mr. Hipkiss. Actually it was Destutt's daughter who married Lafayette's son just prior to Lafayette's letter cited above.

oblige me by appointing the time that will be most convenient to yourself. I have a great desire to exert my abilities in this portrait, and your indulgence will greatly obligate,

Honored Sir your very
Humble Serv
C. W. PEALE

Sir

Your favor of setting today will oblige
Your very Hble Serv
C. W. PEALE
Tuesday morning.

These letters appear in sequence, undated. The letter preceding them is dated Dec. 3, 1791. The next dated letter (written after an illness and an apparently considerable lapse of time) is of January 12, 1792. They establish, as Hart stated on the basis of the inscription to the engraved version, that Peale painted Jefferson in 1791—and establish that this was in December.

Until after 1916, only one Jefferson portrait of Charles Willson Peale was publicly known.[47] The catalogue of the sale of Peale's museum in 1854 lists as no. 82 a Jefferson, without name of artist, but in the group which was there tacitly held forth as the work of Charles Willson Peale.[48] This has been presumed to be the one (fig. 8), preserved at Independence Hall, where it has been since before 1875, if not since 1854. Jefferson's hair is shown in its natural reddish hue, drawn back from the face.

This portrait,[49] characteristically for the Peale Museum series, is framed in an oval with only the ground extending outside this area. The original pigment is lost in many small spots on the left side of the face, and there is much repainting, including the hair and the eyes. Comparison of x-ray photographs with those of indubitable works of Charles Willson Peale and of Rembrandt Peale shows this picture to be an early copy, by a hand other than that of Rembrandt Peale.

The Henry E. Huntington Library and Art Gallery at San Marino, California, has an almost identical likeness (fig. 9)[50] acquired by Mr. Huntington from M. Knoedler and Company in 1924, the bill containing the following statement:

> The portrait was painted in Philadelphia when Peale and Jefferson were both there in about 1776. . . . John Condit of Newark, N. J., bought the portrait, probably at the Peale sale in 1854. He left it to his youngest brother, who gave it to John Porterfield, a descendant of Jefferson.[51]

The date is obviously incorrect, and we have seen that there was no such purchaser of a Jefferson portrait at the sale of 1854. The statement nevertheless has a good basis in fact, as Mrs. William Porterfield, mother of John, kindly informs us. His surviving brother, Wilson, writes as follows:[52]

> The Jefferson portrait was bought by John Conduit of Newark, N. J. who left it to his youngest daughter Tacie. Tacie was a great friend of father's and gave it to John. John Porterfield was born March 7, 1900. The picture was given to him on Christmas of the same year. . . . My brother died July 5, 1916. After John's death, it was sold by Thomas E. Kirby, American Art Association, Madison Square South,

---

[47] The Pennsylvania Academy of the Fine Arts *Catalogue of an Exhibition of Portraits by Charles Willson Peale and James Peale and Rembrandt Peale*, 1923, includes no portraits of Jefferson by any of these artists.

[48] Mr. Charles Coleman Sellers writes us: "I have three marked catalogues of the Museum gallery sale in 1854. In one of these someone, possibly Rembrandt Peale, corrected the authorship of all the pictures. The two portraits of Jefferson, nos. 82 and 152, are both attributed to Rembrandt by this hand. The cataloguer assumed that 82 was by Charles Willson Peale, and attributes 152 specifically to Rembrandt. The purchasers were:

82. Schatt (also spelled Shott) $135
152. P. C. Herbert (also spelled Herben, Herban) $145"

The late Mantle Fielding had a copy of the catalogue in which the buyers were noted by John McAllister, who attended the sale. In the transcript of this, with additional notes of Fielding, at the Frick Art Reference Library, the entries are:

82. Schott (Bryan, N. Y.) $135
152. P. E. Erben (Bryan) $135

Fielding states that "P. E. Erben was probably a dealer, as many of the paintings purchased by him went to the City of Philadelphia (Independence Hall) Mr. Bryan (New-York Historical Society)."

[49] Oil on canvas, 23½ × 19 in. Not signed or dated on the face. Restored and relined in 1919, so that we cannot know whether the back of the original canvas bore any inscription. Mr. Horace T. Carpenter, Curator at Independence Hall, kindly permitted this picture to be removed from its frame for examination and for x-ray photography performed by Mr. Henri Marceau, who has kindly given me the advantage of his highly qualified opinion on this and several other Jefferson portraits.

[50] Oil on canvas 26⅛ × 22¼ in. Relined, with label "Restored at the Copley Gallery, 103 Newbury Street, Boston, May 16, 1919."

[51] Kindly supplied by the Curator, Mr. Maurice Block, who also generously had x-ray negatives taken of the picture for me.

[52] Obligingly communicated by Mr. Virginius R. Shackleford.

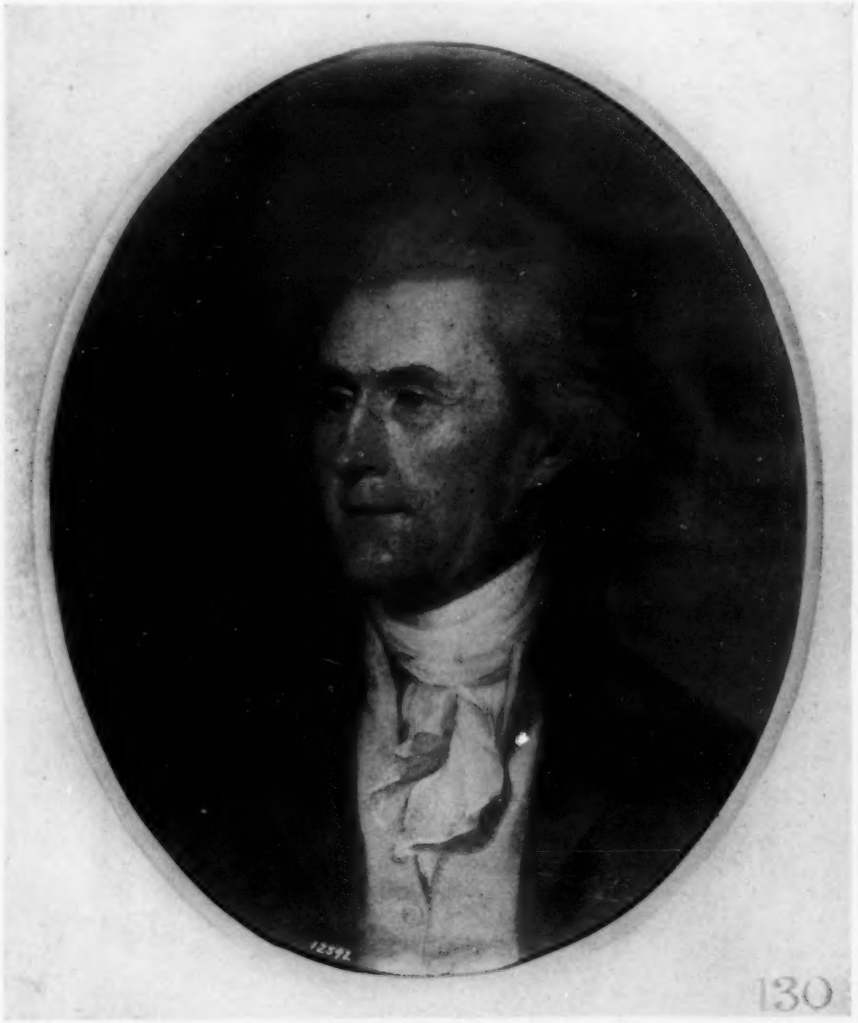


FIG. 8. Jefferson, after Charles Willson Peale, 1791. National Museum, Independence Hall, Philadelphia.

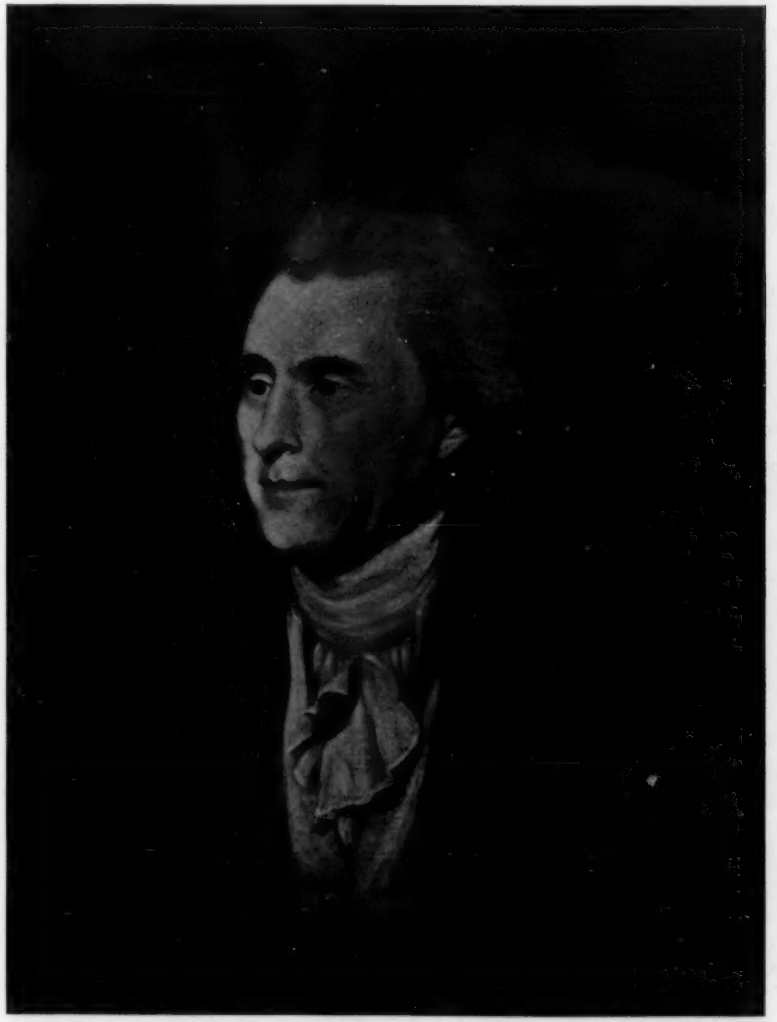


FIG. 9. Jefferson, after Charles Willson Peale, 1791. Henry E. Huntington Library and Art Gallery, San Marino, California.

New York City, to a Mr. Mearus of Cincinnati, Ohio.[53]

In this version the background is carried out to the full size of the rectangular canvas, which was not characteristic of the series of portraits by Charles Willson Peale in his museum. The picture to judge by x-ray negatives,[54] has not been so much restored as the one at Independence Hall.

Careful comparison of radiographs of both pictures with those of undoubted specimens from the hand of Charles Willson Peale was kindly made for me by Mr. Henri Marceau, whose conclusions, in which I fully concur, are as follows:

> The Huntington portrait of Jefferson and the one at Independence Hall appear to be old copies, perhaps from the same lost original. Their radiographs display none of Peale's form-building style, the strokes being indefinite and without function in defining and creating the essential features in the portraits. Likewise the costume elements are vaguely drawn, lacking the spirit of Peale's drawing. These characteristics are those of the copy in which the copyist has attempted to reproduce the painted surface of the original.

The sculptor Giuseppe Ceracchi came to America in 1790, hoping to secure important commissions in the new republic, where an equestrian statue of Washington was proposed. By 1792 he was back in Genoa, having modelled busts of Washington and Jefferson among others. He wrote Jefferson from Munich March 27, 1793, stating that he was going to Leghorn where he expected to find his boxes containing the models he made in America. "I shall," he wrote, "then have the pleasure to cut your bust in marble and present it to you." Jefferson's disclaimer of November 14 he answered from Florence, saying the bust was finished and that he would "ask your permission not to be defrauded of the honour of dedicating it to your family." On May 11, 1795, however, when back in Philadelphia, he sent a bill to Jefferson for the bust. While this dubious proceeding, which the artist employed in certain other cases also, was followed by a considerable delay, Jefferson must certainly have admired the work, for on June 10, 1797, he noted in his accounts: "Agreed with . . . Bohlen to give 300 livres tournois [about $60] for my bust by Ciracchi if he shall agree to take that sum," and on January 14, 1800, he noted: "Gave J. Bohlen order on J. Barnes for 550 Dol. paiment for a marble bust of myself by Ciracchi."

According to the indications of Cornelia Jefferson Randolph, Jefferson's granddaughter, the Ceracchi bust stood in the hall at Monticello opposite the bust of Hamilton—"opposed in death as in life," as Jefferson is quoted by his biographer Rayner as saying to visitors. It was still at Monticello on the day of a sale of Jefferson's effects, January 17, 1827, and those present signed a petition urging its acquisition by the State, as "the only sculptured memorial in Virginia of the Author of the Declaration of Independence."[55] It passed, however, to the United States, and was destroyed in the fire in the Library of Congress in 1851.[56]

Of the Ceracchi bust a cast (figs. 10 and 11) survives at Monticello, when taken we know not.[57] Today we prefer the vivacious realism of Houdon to the classic idealism or Roman characterizations of Ceracchi, but we can by no means be sure that this preference was shared by the men of 1800. Dr. William Thornton, the gifted amateur architect, who himself twice copied Stuart's profile of Jefferson, and who knew him intimately, wrote that Ceracchi's was a "superb bust, one of the finest I have ever beheld."[58]

On July 12, 1792, in Philadelphia, Jefferson noted: "Pd Williams for drawing my portrait 14 D." This would presumably have been William Joseph Williams,[59] who made the rather forbidding masonic pastel portrait of Washing-

---

[53] In spite of this circumstantial statement, I have not succeeded in identifying the sale or catalogue number, so as to fill in the pedigree between this purchaser and Knoedler's.

[54] Unfortunately when the picture was relined the adhesive used was white lead, which by its relative opacity and uneven application obscures the x-ray negatives in numerous places.

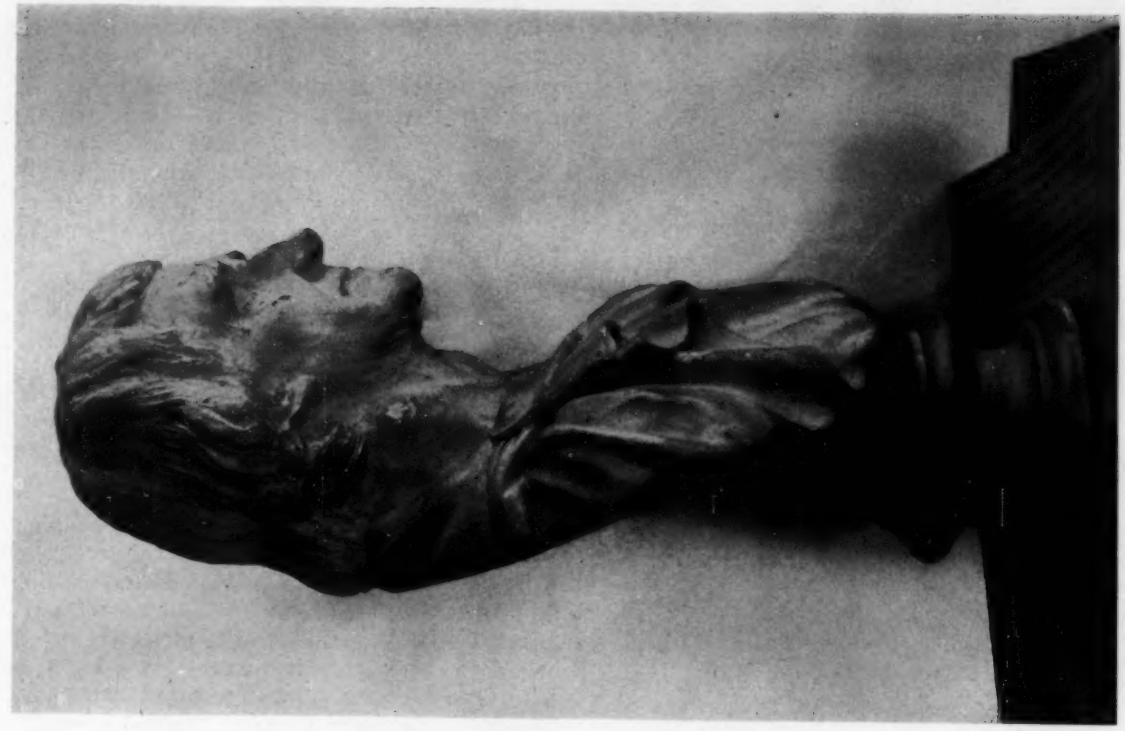
[55] Manuscript published in *Tyler's Quarterly* **8**: 243–246, 1927.

[56] *Washington Intelligencer*, December 25, 1851.

[57] Neither surviving members of the Randolph family of Edgehill and Lego, of whom I have inquired, nor Mrs. Charles Mayhoff, the surviving sister of the late Jefferson M. Levy, former owner of Monticello, were able to inform me. Mrs. Mayhoff states that the bust was already there in the time of her uncle, Commodore Uriah P. Levy, who acquired the place in 1836 and died in 1862.

[58] Quoted by John Hill Morgan, *Virginia Historical Portraiture*, 253. I judge this is from Thornton's letter to Jefferson, May 27, 1817, Library of Congress (Jefferson papers currently inaccessible).

[59] The date is surely too late for the other William Williams, active 1746–1775. *Cf.* W. Sawitzky, William Williams, First Instructor of Benjamin West, *Antiques* **31**: 240, 1937; and Further Light on William Williams, *New-York Hist. Soc. Quarterly Bull.* **25**: 101–112, 1941.



FIGS. 10 AND 11. Jefferson, by Giuseppi Ceracchi, 1790 or 1791. Unique plaster from the lost marble which belonged to Jefferson. Thomas Jefferson Memorial Foundation, Monticello, Virginia.

ton. Of his Jefferson, the small value of which may well be reflected in the low price, we hear and know nothing further.

To one who, like myself, knew and admired the late Lawrence Park, and who, as everyone must do, appreciates his great service to scholarship through his monumental work on Gilbert Stuart, it is painful to have to remark that the account of the Stuart Jeffersons in his book contains a number of lamentable errors.[59a] Such statements as that "Mrs. Jefferson considered [the medallion portrait] the better of the two original portraits of her husband by Stuart,"[60] whereas Jefferson's wife died a decade before Stuart came back to settle in America, can only be explained by a general unfamiliarity with matters Jeffersonian which Park's untimely death did not give him time to repair. This unfamiliarity extended to the Jefferson documents and letters, knowledge of which would have precluded several more serious misstatements both as to identities and dates of the portraits.

Jefferson's pocket account books list the following payments to Stuart:

1800 May 12. paid Stuart for my portrait 100 D.
1805 June 7. pd. Gilbert Stewart for drawing my portrait 100 D.

As to the first of these payments, several points may be noted. Jefferson was in Philadelphia. May 12 was two days before the adjournment of Congress, where he presided over the Senate. The payment to Stuart was one of no less than fourteen payments by Jefferson on that day, totalling over $750—out of the proceeds of a note for $1250 discounted on the 10th—settling various accounts preparatory to his departure for home on the 15th. We are accordingly not to assume necessarily that Jefferson had completed his sittings that very day. On the other hand, we have no reason to believe, as Park and others have supposed, that Jefferson's first sittings were in 1799. Aside from Jefferson's own subsequent statement that he sat to Stuart in May 1800, we may remark that it was a period when he was well in funds, was spending money freely, and paying promptly where, as here, there was no dispute. There would thus have been no occasion for any substantial delay in his paying Stuart.

Regarding Jefferson's second payment to Stuart which, as we shall see, was for the medallion head, Jefferson wrote Stuart June 18, 1805:

> Mr. Jefferson presents his compliments to Mr. Stewart, and begs leave to send him the enclosed for the trouble he gave him in taking the head à la antique. Mr. Stewart seemed to contemplate having an engraving made either from that or the first portrait; he is free to use either the one or the other at his choice; the one not proposed to be used I will be glad to receive at Mr. Stewart's convenience; the other when he shall be done with it.[61]

From later letters we shall see that these payments were made at the time of the first and third of three life portraits, the second having been painted to replace the first, because Stuart expressed himself as dissatisfied with that. Jefferson wrote to Joseph Delaplaine, May 30, 1813:

> You enquire whether there is an approved portrait of myself, by whom painted and in whose possession. Mr. Stuart has drawn two portraits of me, at different sittings, of which he prefers the last. Both are in his possession. He also drew a third in water colours, a profile in the medallion stile, which is in my possession. . . .

With the added excuse of helping Delaplaine, Jefferson now wrote to Stuart August 9, 1814:

> You wished to retain the portrait which you were so kind as to make of me while in Washington, until you should have time to have a print copied from it. This, I believe has been done, at least I think I have seen one which appeared to have been taken from that portrait. Mr. Delaplaine of Philadelphia is now engaged in a work relating to the general history of America, and, wishing it to be accompanied with prints, has asked permission to have one taken from the same original, adapted to the size of his volume. I have therefore authorized him to ask for the portrait in your possession, to copy his print from it, and return it to me. This I have done on the supposition that I had not been mistaken in believing it had already answered in your hands the purpose you had meditated.

[59a] The following discussion of the Stuart Jefferson was first embodied in the writer's paper "Gilbert Stuart's Portraits of Jefferson," *Gazette des Beaux-Arts*, **6**, ser. 26: 95–112, 1944.

[60] Park, *Gilbert Stuart* I: 440, Rudge, 1926. This rests on a statement by Mrs. William B. Harrison, quoted in the *History of the Centennial Celebration of the Inauguration of George Washington*, 485: "My grandmother [i.e., Martha Jefferson Randolph, Jefferson's daughter] thought it the better likeness of the two hanging at 'Monticello.' "

[61] Quoted from Hart in *McClure's Magazine*, *loc. cit.*, 48. I have not seen the original of this note to verify whether the editors have done injustice to Jefferson in transcribing the French phrase.

This not being effective, Jefferson in 1818 turned for help in Boston to Henry Dearborn, his former Secretary of War. A letter of July 5, 1819, in correspondence from which we shall quote further, in his fullest summary of the case:[62]

> With respect to Mr. Stuart, it was in May, 1800 I got him to draw my picture, and immediately paid him his price, one hundred dollars. He was yet to put the last hand on it, so it was left with him. When he came to Washington in 1805 he told me he was not satisfied with it, and therefore begged me to sit again, and he drew another which he was to deliver to me instead of the first, but begged permission to keep it until he could get an engraving from it. I soon after got him to sketch me in the medallion form, which he did on paper with crayons. Although a slight thing I gave him another 100 dollars, probably the treble of what he would have asked. This I have; it is a very fine thing, though very perishable.

We thus observe that Stuart's second life portrait must have been painted some time prior to June 7, 1805, the date of Jefferson's payment for the medallion portrait which followed it.

What was the character of the first portrait of Jefferson painted by Stuart in 1800? Park supposed it to be the one at Bowdoin College, but that, as we shall see, was not commissioned and painted until 1805. The other known Jefferson portraits by Stuart—except the profile medallion, which was also painted in 1805—show heads of identical pose and general character with the Bowdoin head, and represent the type created in that year.

Strangely enough, no one seems to have thought hitherto of looking among the engraved portraits of Jefferson for the clue to Stuart's first portrait of him, but there it is readily to be found. Stuart advertised in the *Philadelphia Aurora* of June 12, 1800, under the heading,

> ADAMS AND JEFFERSON:
>
> Mr. Stuart informs the public that engravings from his Portraits of the President and Vice-President are likewise preparing under his immediate direction, and will be published in a few weeks.[63]

It has generally been supposed that this project was not directly carried into execution, and indeed no American engraving which fulfills it is known to me.[64] There are, however, a group of engravings after Stuart, prior to 1805, which can be derived only from his portrait of 1800.

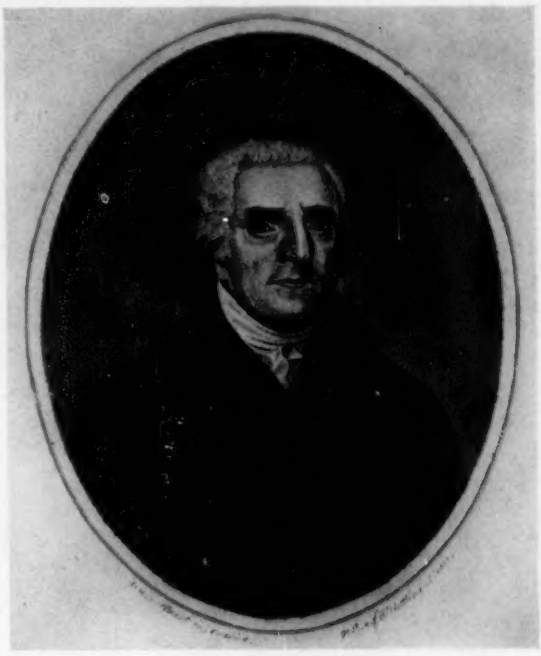


FIG. 12. Engraving after Stuart's first life portrait of Jefferson, painted in 1800. Engraved by W. Nettling. Photograph by courtesy of the University of Virginia.

Of these the earliest I know[65] appears to be the stipple inscribed: "Thos. Jefferson President of the United States of America. Painted by Stuart in America. Sold and published August 1, 1801 by Edw. Orme, New Bond Street, London." This is clearly the source of one (fig. 12), with identical wording, translated into German "gest. v. f. W. Nettling, L. 1801."[66] The same type, again with English wording, reappeared in the *European Magazine* "Published by T. Sewell, Cornhill, June 1, 1802."[67] All these are of full bust length, nearly full face, with hair drawn back from the forehead as worn by Jefferson hitherto, with very prominent nose, body fron-

[62] *The Writings of Thomas Jefferson*, ed. by A. E. Bergh and A. A. Lipscomb, XIX: 271, 1907.

[63] Published by John Hill Morgan in his biographical sketch in Lawrence Park's *Gilbert Stuart* I: note 60, Rudge, 1926.

[64] None is listed in Stauffer's or Fielding's works on early American engravings.

[65] Oval, $4\frac{3}{16} \times 3\frac{3}{16}$ in., listed in the catalogue of the Hampton L. Carson Collection sold in 1904, No. 1165. Kindly called to my attention by Carl Zigrosser.

[66] Oval, $4\frac{1}{8} \times 3\frac{3}{16}$ in. Carson Collection No. 1194. An example owned by the University of Virginia to the Jefferson Bicentennial Exhibition at the National Gallery.

[67] Carson Collection No. 1166. An example was lent to the Exhibition by Mr. Hall Park McCullough. There are still further derivatives of this type, undated.

tal, the vest buttoned rather high, with little of the neckcloth visible. As reflected in these engravings, the portrait can have been neither very pleasing nor very characteristic of the subject.

We see that when Stuart told Jefferson he was not satisfied with the portrait of 1800 he was telling the truth. No such oil portrait as the one shown in these engravings is known to exist, nor to have existed since its successors were painted in 1805. We must conclude beyond peradventure that Stuart "rubbed it out," as he said he did one of his life portraits of Washington.[68]

The other letters which passed between Jefferson and Dearborn, still unpublished, are very illuminating, and of interest also as to Stuart's personality and manner of dealing:

Jefferson to Dearborn, October 27, 1818:[69]

Can you without involving yourself in offense with Stewart obtain thro' any channel a frank and explicit declaration on what ground he detains my portrait? In what term? And whether there is to be an end of it. I think he has now had it 10 or 12 years. I wrote to him once respecting it, but he never noticed my letter.

Dearborn to Jefferson, November 16, 1818:

As there has been a much greater intimacy between my Son and Stewart than between Stewart & myself I requested my son to call on him and endeavor to obtain such frank & explicit information from him as you desire. An interview took place and after many trifling excuses for the long detention of the portrait and its unfinished situation, he said that he could not finish it, in cold weather but would certainly complete it in the Spring. We will endeavor to push him on and if possible compel him to finish it as early in the Spring as we can.

Dearborn to Jefferson, June 24, 1819:

Having not yet been able to prevail on Stuart to finish your portrait I suspect that you have paid him in part or in full in advance if so I should like to know it, as I might in that case address his pride with some change of success. If you have not made any advance and will authorize me to pay him as soon as he shall complete it I will address his poverty which is now great and by engaging to pay him and by frequent calls I should hope to succeed.

There follows the letter of Jefferson to Dearborn July 5, 1819, previously quoted. Dearborn replies January 20, 1820:

After frequent promises Mr. Stuart has again forfeited his engagement to finish your Portrait. His last promise was made in October last when he said he would have it done by the first of January but on calling on him I found he had not touched it. Feeling a little out of patience I observed to him that I would inform you that you must never expect to have it. I then indicated his having received pay for it. He said that you paid him an hundred dollars for one that you now have in your home and that he received one hundred dollars for a Medallion, but had received nothing for the one he now has. That he painted this for himself. That he had no commission from any one to paint it. I was too much out of temper to say anything more to him and retired.

Jefferson to Dearborn, February 5, 1820:

On the subject of Mr. Stewart and my portrait, he must have spoken without reflexion, when he supposed it in my possession and hanging in my hall. The peculiarities of his temper and ideas render him a difficult subject to handle. In the enclosed letter I have endeavored to bring his recollection to rights as softly as I can. With respect to the 1st canvas portrait I thought it a good one, and should have been contented with it, had he not himself been dissatisfied with it, and still if he chuses to deliver that instead of the 2nd if he will finish and deliver it I shall be satisfied. I am sorry to give you this trouble, but if this letter does not effect a delivery the business may still be awhile longer.

Dearborn to Jefferson, March 3, 1820:

On receiving your letter of the 5th of Feby enclosing one to Stewart I requested my son to call on Stewart and give him your letter and hear what he had to say. He now owns that he had been mistaken and that he has received one hundred dollars for the portrait, which you have not received and only wants to know whether you would prefer a common portrait or one of half the length of the Body, the former at $100, the later $300. He said he would write to you on the subject but he may neglect it as he has the finishing of the picture. If you will inform me which of the two you would prefer I will inform him.

Jefferson to Dearborn, March 26, 1820:

The negotiation with Mr. Stuart has given you much more trouble than I had expected and more than it should have given had I expected it. However we may now hope to close it by accepting one of the alternatives he proposes. I shall be perfectly content to receive the original he drew in Philadel-

[68] *Cf.* John H. Morgan and Mantle Fielding, *The Life Portraits of Washington*, 228, Phila., 1931.

[69] As the originals belonging to the Library of Congress are currently inaccessible, we quote the correspondence from the transcripts by Charles Henry Hart, in the Frick Art Reference Library.

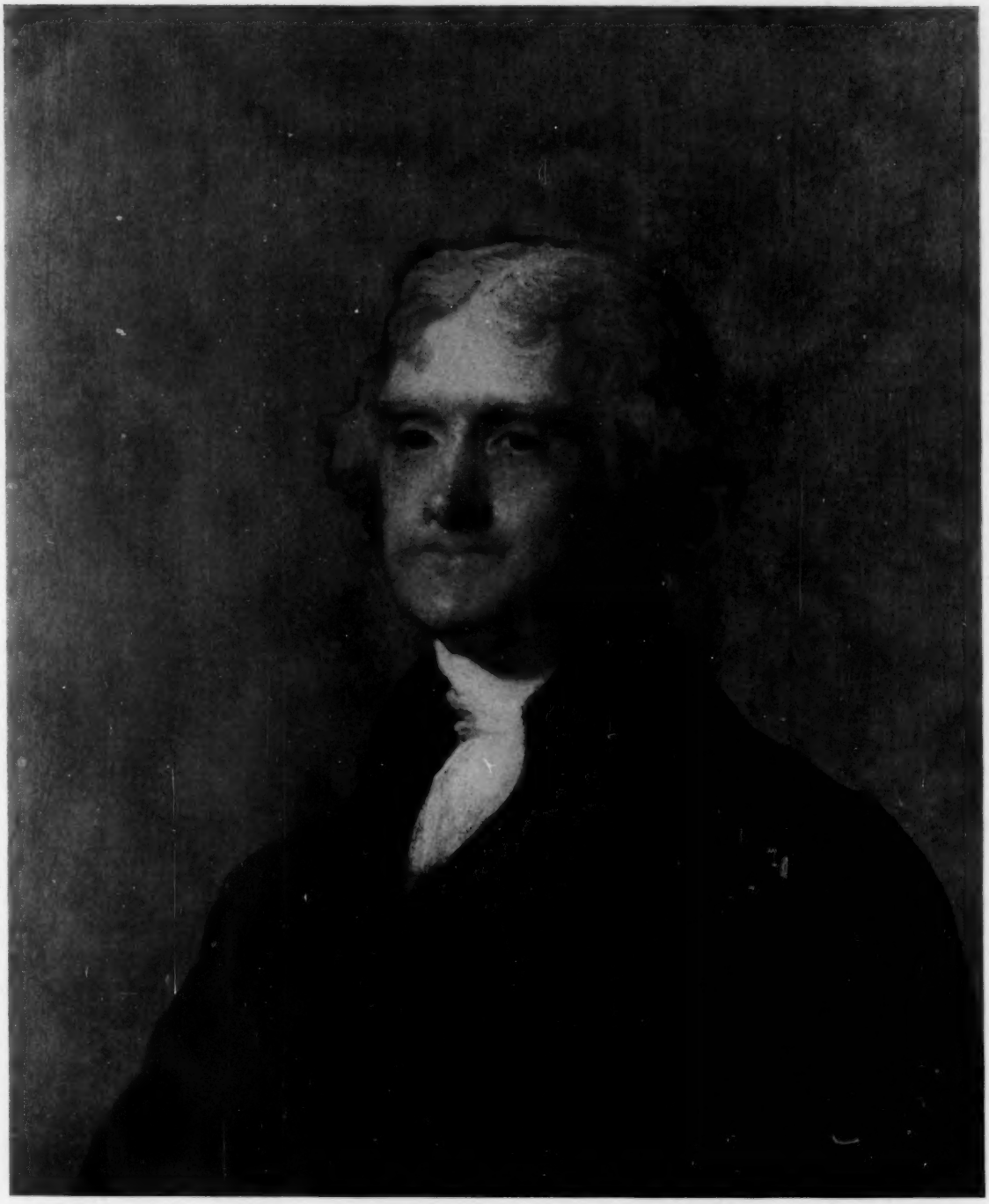


FIG. 13. Jefferson, life portrait by Gilbert Stuart, 1805. Estate of Percy Selden Straus. Photograph by courtesy of The National Gallery of Art.

phia in 1800, which was of the common size (that the painters call I believe a bust). It will suit me better than a half length as it will range better in the line of my other portraits not one of which is halflength I have no doubt that Mr. Stuart's justice will think me entitled to the original and not merely a copy. There was something pleasanter in the aspect of that portrait which I liked better than the second drawn at Washington. It will come safest by water addressed to the care of Capt. Bernard Peyton, Richmond.

Dearborn Jr. to Jefferson, May 21, 1821.

I have procured your picture from Mr. Stuart at last and ship'd it on board the brig Richmond, S. Hill Master for Richmond to the care of Capt. B. Peyton.

On August 17, 1821, Jefferson could thus finally write to Dearborn:

The portrait by Stuart was received in due time and good order and claims, for this difficult acquisition, the thanks of the family.

This is the Stuart Jefferson known as the Edgehill portrait (fig. 13)[70] from its descending to the Jefferson Randolphs of Edgehill. It remained in the family by inheritance until about 1902, when it was purchased by a collateral descendant, the Honorable Francis Burton Harrison. Acquired by the Babcock Galleries in 1927, it was first sold to the Honorable John G. Winant, and subsequently to the late Percy S. Straus.

Jefferson is shown three-quarters to left, with white neckcloth not largely showing, with locks of hair now falling on his forehead—a change of coiffure since any of the earlier portraits, except the Saint-Mémin of the previous years. The face has very rich carnations, with scumblings in blue. The background is little more than a ground of bluish white, through which the wood panel may be seen. The work is obviously a life portrait: above the head we see where Stuart twice felt for the contour before arriving at the final outline.

Though Jefferson, in desperation, had finally expressed himself willing to accept even the life portrait of 1800, rather than be sandbagged into paying extra for a half-length, he insisted on a life portrait; the life portrait of 1805 was what he had expected and wanted, and that was what he in the end received. We learn that it had still been unfinished in 1818; at that time, like the Athenaeum Washington, it may well have comprised little more than the head. As delivered, it still has the neckcloth merely indicated, without the frills added by Stuart in later versions; it also lacks the ribbon there added to the queue. It has all the immediacy of a work from life. The hair is almost entirely white; apparently Jefferson powdered it again at this period, for his natural hair retained much of its color long after this. Jefferson's face here is somewhat broader than in replicas made from this portrait; his expression has the mildness and benignancy remarked in all descriptions of him in his later years.

Jefferson's great-granddaughter, Mrs. William B. Harrison of Edgehill, stated about 1892[71] that when delivered

. . . the paint on the coat of the full-face portrait was fresh, and Mrs. Randolph [Jefferson's daughter] thought Stuart had sent a replica in place of the original. Stuart heard of the suspicion and on meeting my grandmother in Philadelphia expressed his mortification at her supposing him guilty of such a thing, and assured her he had sent the original, telling her he often left the accessories to be filled in just before it was sent home.

Actually Mrs. Randolph can never have met Stuart after the delivery of the picture, for he was not south of Boston after 1805 and she did not go there until 1833, after his death in 1828. If there was any such conversation with Stuart at all, it can only have been with her son, Mrs. Harrison's father, Thomas Jefferson Randolph, who was in Boston in 1825.

Writing to Joseph Delaplaine on May 3, 1814, Jefferson spoke, as we have seen, of the "two original portraits of myself taken by Stuart," as then still "both in his possession in Boston," and said: "The President has a copy from that which Stuart considered the best of the two," i.e., the second, of 1805. Again on May 20, 1816, he wrote to Delaplaine: "I have no doubt the copy of my portrait at the President's residence in Orange . . . is as good as the original in the possession of Stuart." This picture, also a head and shoulders (fig. 14)[72] passed by sale, on March

[70] Oil on panel, 26½ × 21¾ in.

[71] As quoted in the *History of the Centennial Celebration of the Inauguration of George Washington*, ed. by C. W. Bowen, 484, 1892.

[72] Oil on canvas, 30 × 25 in., the size, greater than that of the Edgehill Jefferson, doubtless being adopted to conform to that of Madison's own portrait painted by Stuart in 1804. Mrs. W. B. Harrison, in the statement already cited, says she made it because "Governor Coles maintained that his was the original and ours the replica."

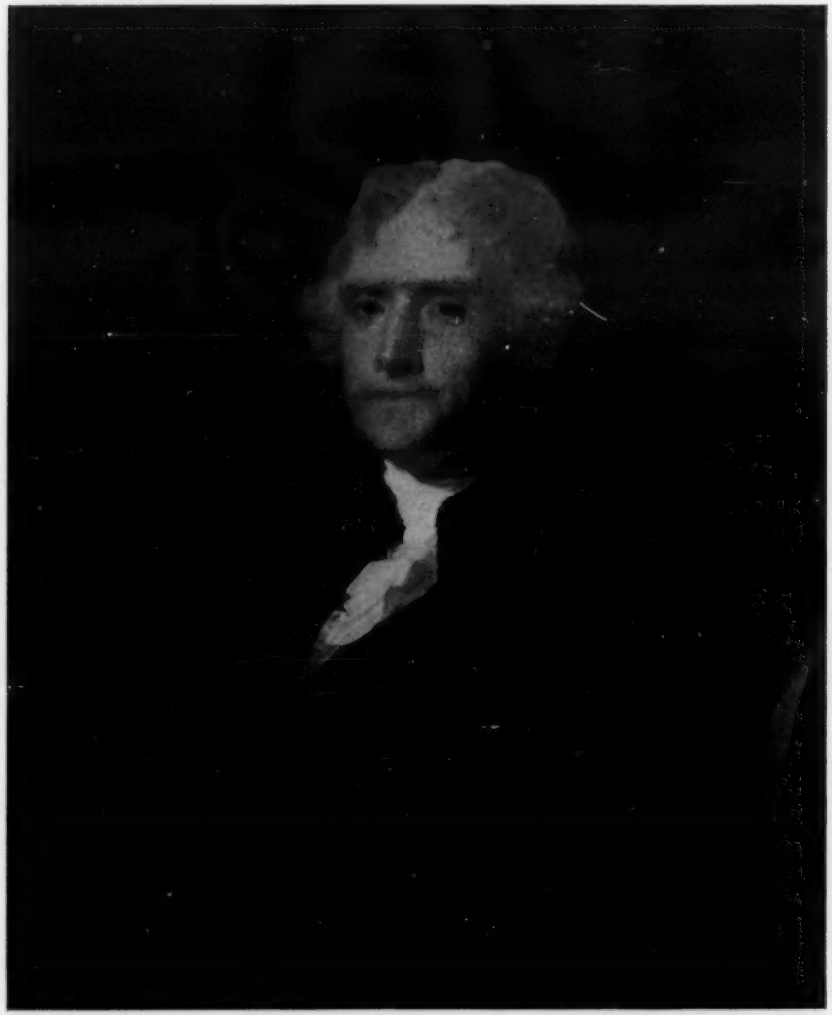


FIG. 14. Stuart's first replica of the second Jefferson life portrait, painted for James Madison. Private collection. Photograph by courtesy of The Philadelphia Museum of Art.

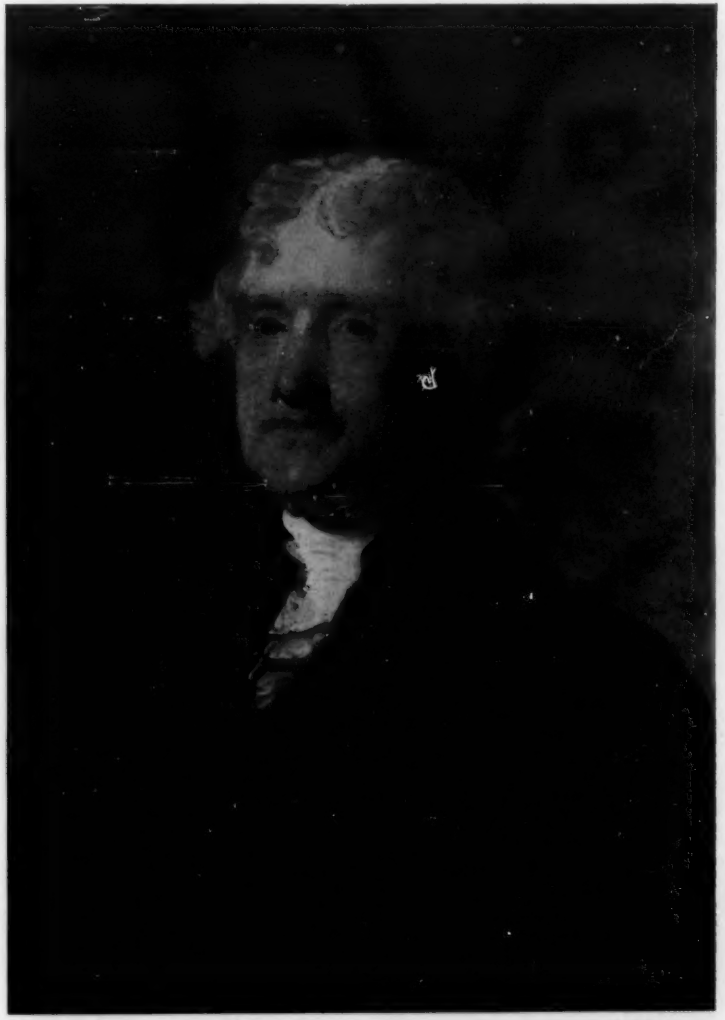


FIG. 15. Replica painted by Stuart for George Gibbs about 1825 of the type of the second Jefferson life portrait. Honorable T. Jefferson Coolidge. Photograph by courtesy of the National Gallery of Art.

1, 1851, after the death of Mrs. Madison to Judge Edward B. Coles,[73] then descended to his son Governor Edward Coles and to his daughter Miss Mary Roberts Coles, by whom it was bequeathed in 1941 to her nephews.

Claims that another picture than either of these is the original life portrait by Stuart, whether from the first or the second sitting, have been made also for the seated half-length painted for James Bowdoin, in the collection of Bowdoin College (fig. 16).[74] Park described this portrait as painted in Philadelphia in 1799, and stated that "it was painted from life."[75]

The present Director of the Bowdoin College Museum of Fine Arts, Mr. Philip C. Beam, has very generously supplied the following documents regarding its genesis:

James Bowdoin in Boston to Henry Dearborn in Washington, March 25, 1805:[76]

. . . I shall be much obliged to you to procure me the portraits of Mr. Jefferson and Mr. Madison if a good painter can be found in Washington, and they sh[d]. be willing to take the trouble of sitting therefor. I sh[d]. like to have them done by Stuart, could he be induced to execute them, as well he is able. They need not be framed, as I can procure more fashionable frames and better frames in Europe. Please to let y[e] Pictures be half length and of a size to match each other . . .

Dearborn in Washington to Thomas L. Winthrop in Boston, June 27, 1805:[77]

By Mr. Bowdoins request I ingaged Mr. Stuart to take a half length portrait of the President of the U. S. and one of Mr. Madison. Mr. Stuart has nearly completed them, and will take them with his other effects to Boston, and when completed there will deliver them to you, to be forwarded to Mr. Bowdoin, and as Mr. Bowdoin requested me to draw on you for the expence of the two portraits, I take the liberty of requesting you to pay to Mr. Stuart the amount of his bills when presented.

Winthrop to Bowdoin abroad, August 14, 1807:[78]

. . . The pictures of Mr. Jefferson and Mr. Madison remain with Mr. Stuart; you have omitted to give any directions respecting them.

Bowdoin in Paris to Winthrop, October 13, 1807:[79]

. . . With respect to Mr. Jefferson's and Mr. Madison's pictures, I wish them to be retained to be put up in my house. . . .

This correspondence establishes that actually the Bowdoin picture was painted on a half-length canvas in Washington in 1805, that it was nearly completed by June 27—twenty days after Jefferson paid Stuart for the third life portrait—and that it was not delivered before 1807 at earliest—as we shall see also from other evidence. Bowdoin did not return from Europe until the spring 1809, and died in 1811.[80]

That the Bowdoin Jefferson was indeed not finished until 1807 is shown by a factor which, strangely enough, has not been previously considered in connection with any of Stuart's works. The chairs in which he shows his sitters, in the pose he so often repeated, vary significantly with the places and times at which they were executed. Thus, for instance, a very considerable group[81] show a Directoire bergère with arm having a lyre-like outward scroll, delicately carved on the face. We find that every single one of the pictures in this group was painted in Boston, those specifically datable being from the years 1808, 1809, 1810 and 1811. Similar instances of other types of chairs used for restricted periods and, for example, only in Philadelphia, could be multiplied. Stuart, for variety, often used more than one type of chair at the same time, but no given type

[73] Information kindly supplied by Mr. William Sawitzky, citing the advertisement in the Washington *National Intelligencer* of that date.

[74] 46¼ in. high by 38¼ in. wide.

[75] John Hill Morgan concurred in this opinion, in *Virginia Historical Portraiture*, 251, 1930. The source of most current statements regarding the picture is *The Descriptive Catalogue of the Bowdoin College Art Collections*, by Henry Johnson, Curator, 1895, which lists it as No. 83, with the note: "This portrait and the following [of Madison] were painted from life for Mr. Bowdoin . . . who was the personal friend of both Presidents." In the Historical Introduction, page 4, the author states: "About the year 1821 Mr. Gilbert Stuart visited the collection more than once for the purpose of copying his paintings of Presidents Jefferson and Madison"—concerning which we shall have more to say below. Some of Stuart's opinions on the pictures in the Bowdoin collection are quoted in this catalogue.

[76] James Bowdoin III's letter books, in the Library of Bowdoin College.

[77] Letter on file in the Bowdoin Museum of Fine Arts.

[78] Letter on file in the Bowdoin Museum of Fine Arts.

[79] James Bowdoin III's letter books, *loc. cit.*

[80] The paintings remained in the hands of his widow until 1813, when she married Henry Dearborn and they were delivered to the College.

[81] E.g., Park Nos. 41, 122, 134, 212, 214, 215, 715, 774. The chair arm is of a type just then adopted in America in the work of such craftsmen as Duncan Phyfe and Samuel McIntire.

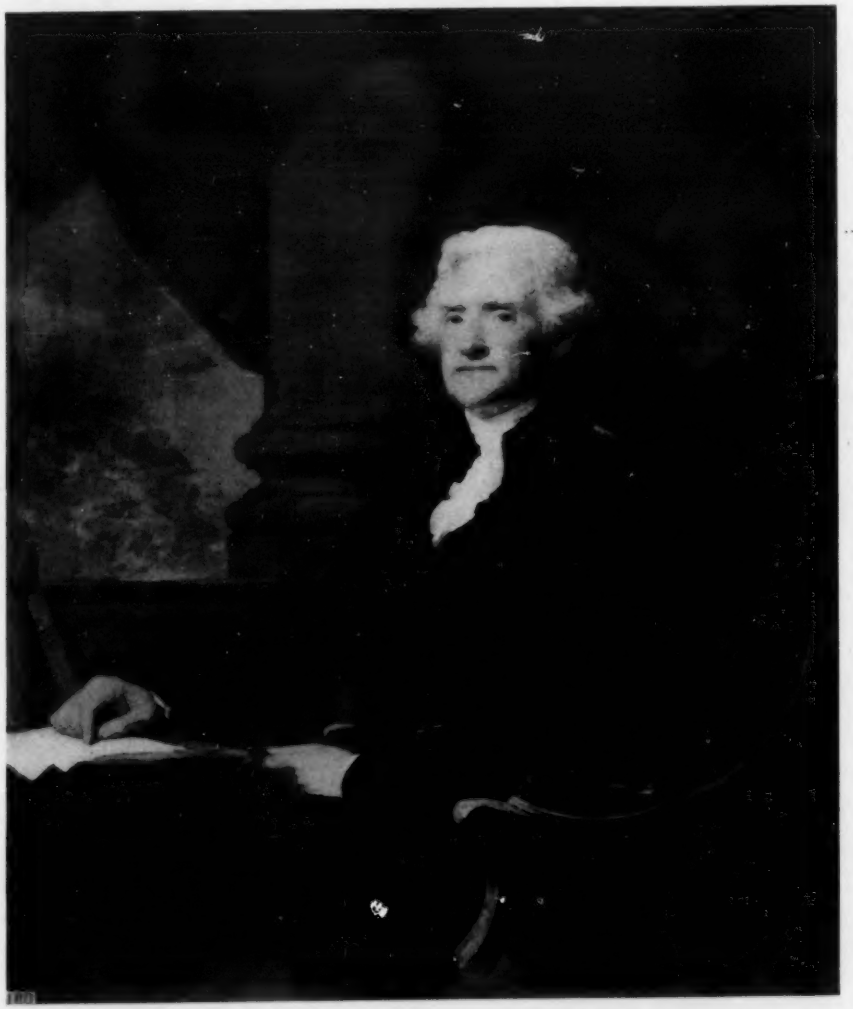


FIG. 16. Stuart's portrait of Jefferson painted for James Bowdoin, 1805–1807. The head a replica of the second life portrait. Collection of the Bowdoin College Museum of Fine Arts. Courtesy of Bowdoin College Museum of Fine Arts.

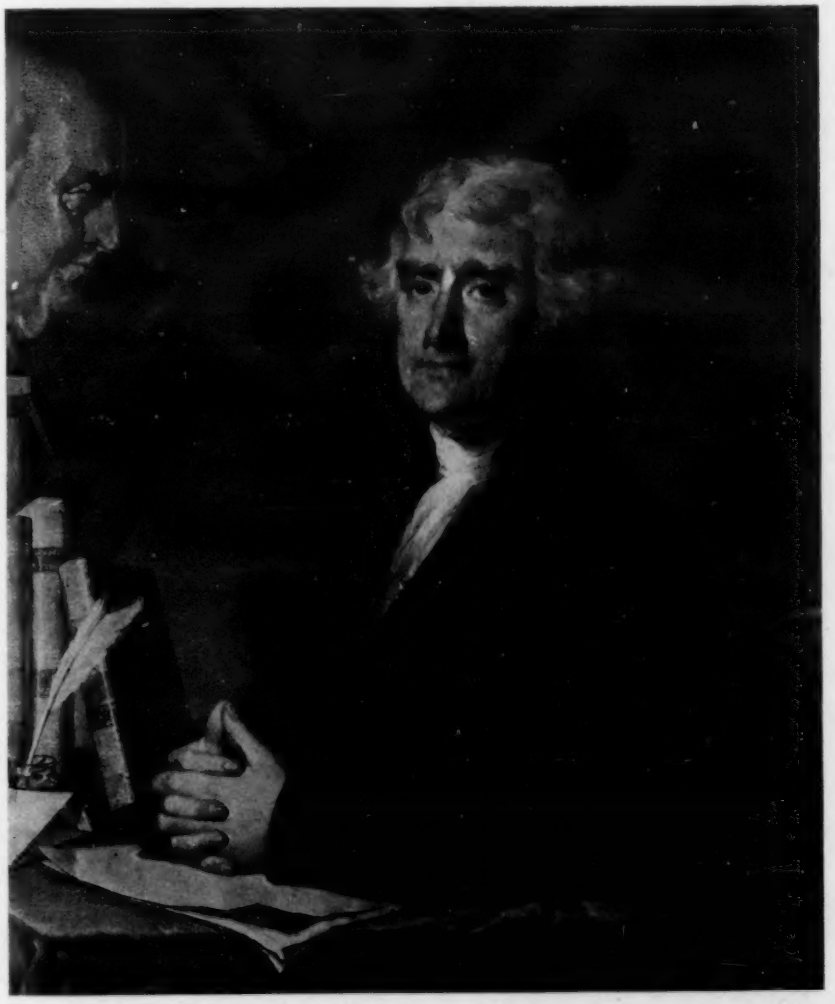


FIG. 17. Stuart's portrait of Jefferson painted for John Doggett, 1823. The head a replica of the second life portrait. Lithograph by Antoine or Nicolas Maurin. Photograph by courtesy of The American Antiquarian Society.

carries throughout his career in different cities. Now the particular arm chair in which Jefferson is shown in the Bowdoin portrait is one which never appears during Stuart's work in New York, in Philadelphia, or in Washington. On the contrary, it is identical with those used in the Boston portraits of Governor James Sullivan, 1807, David Humphreys, after December 1807, and others later. At least the accessories of the Bowdoin Jefferson were accordingly executed in Boston about 1807.

In the case of the Madison there can be no doubt that Bowdoin did not receive a life portrait; the picture, a seated half-length uniform with the Jefferson, follows the head and shoulders painted for Madison himself in 1804. We must conclude that in the case of the Jefferson, where the dates of execution were closer together, Stuart also merely repeated the head and shoulders and later added the accessories. This view indeed is confirmed by an early note of Professor Packard of Bowdoin College, interlined in a manuscript catalogue of 1821, referring to the Madison and Jefferson portraits:

> Stuart declared that he regarded them as good as originals.
>
> A. S. P.[82]

Such a conclusion is entirely in consonance with what we know of Stuart's methods, and with the other evidence. As in the case of the Athanaeum Washingtons, it would have been characteristic of Stuart, above all in a portrait of which replicas would be in demand, to have painted merely the head and shoulders, to keep the original, and to multiply it from time to time at such a size and with such accessories as purchasers desired. This was the case also with the Monroe. The hands are not the hands of Jefferson—they are merely hands. Jefferson had broken his right wrist when riding in Paris in 1786; it was poorly set and greatly enlarged.[83] There is nothing of this in the portrait painted for Bowdoin. It is not a life portrait, but it is a superb picture. It is indeed the equivalent of many of Stuart's replicas of his Washington portraits—replicas in which the subject was idealized in Olympian calm and benignity.

By contrast to the Edgehill portrait, the Madison and Bowdoin Jeffersons—in which the heads are very closely similar to one another—are more "noble," more idealized. In them the accessories are freely developed to a finish.

Two other replicas of the Jefferson portrait of this type were painted by Stuart as part of sets of the first five presidents—sets which must therefore fall after the election of Monroe in 1816. One at half length was for John Doggett;[84] the other, head and shoulders, for Colonel George Gibbs (fig. 15).[85] The latter was sold by Colonel Gibbs' widow to Jefferson's great-grandson Thomas Jefferson Coolidge, and now belongs to his grandson and namesake. The Doggett set, while offered for sale to the nation, suffered in the fire at the Library of Congress in 1851, the Jefferson being among the three of the portraits which were destroyed. It is known, however, through the fine lithograph by Maurin (fig. 17) made from the original, sent to France for the purpose, in 1825.

The lost picture of Jefferson painted for Doggett shows analogies to the Bowdoin Jefferson. Again it is a seated half-length, again with the head and shoulders adapted from the life portrait, again with a table and hands supplied—though this time with hands folded and with additional, different accessories.

The belief exists at Bowdoin College that Stuart made his later replicas of the Jefferson and Madison at Brunswick from the painting there. This has rested on two grounds. First there are the notes of Professor Packard, interlined in the manuscript catalogue of 1821, reporting dicta of Stuart, of which one has already been quoted. Another is:

> I heard Stuart say [of the Jean de Montfort] that he recognized it as [from?] an original by Vandyke. . . . This was said in 1821 or about that time.
>
> A. S. P.

Secondly there is a passage in Professor Sewell's catalogue of the collection (1870):

> After the production of these portraits of Jefferson and Madison Mr. Stuart never again painted from

[82] Kindly communicated by Mr. Beam.

[83] In later life, after he had broken the other arm, and suffered also from arthritis, the size of Jefferson's wrists was described by Webster in 1824 as "extraordinary." H. S. Randall, *The Life of Thomas Jefferson* III: 505, 1858.

[84] Park supposed the set for Doggett to have been painted about 1815. Mabel M. Swan, a close student of Doggett, gives the date as 1822 in her book *The Athenaeum Gallery*, 1940. *Cf.* also her paper, "The 'American Kings,'" *Antiques*, 19: 278–281, 1931. The Gibbs set Park dates as "about 1825." The Doggett picture, to judge by others of the set, was *ca.* 40 × 32 in., the Gibbs-Coolidge picture, 26 × 21⅜ in.

[85] Oil on canvas, 25⅝ × 21⅛ in.

the originals, but copied these, and visited this collection once at least for this purpose.

Mr. Beam informs me that Professor Packard was still alive and active at this time, so that he may well have been Professor Sewell's authority for this statement. Good ground for belief in its truth is the fact that, though Stuart still had the Edgehill Jefferson until the summer of 1821, no Madison was available to him in Boston after 1813.[86] The versions made at Bowdoin would have been the Doggett pictures. The Gibbs-Coolidge set could have been made from these or, if painted while the Doggett set was abroad, would have required another trip to Bowdoin.

What has passed, and was accepted by Park, doubtless on the statement of Bowen as a further replica by Stuart of his Jefferson hangs in the White House.[87] Its provenance as stated by Park is as follows: "It was painted for Lewis Saunders [actually for Col. Lewis Sanders] or at any rate he was the original owner. He sold it to Major William S. Dallam of Lexington, Kentucky, and at his death it was inherited by his daughter who, in 1874, sold the picture to Congress for $1000." We consider this picture, which is very inferior to those we have discussed, to be a copy by or after Matthew Harris Jouett, the "old master of the Bluegrass." Jouett studied with Stuart while Stuart still had the Edgehill Jefferson, and made a copy of it, ultimately multiplying this in several examples which survive,[88] identical in character with the White House picture. To mention nothing else, in this, as in all other Jouett copies of which illustrations are accessible to me, Jefferson's right lapel has a conspicuous notch not found in any of the Stuarts which are fully authenticated.

Stuart's profile medallion portrait of Jefferson (fig. 18)[89] in oils, has always been owned by the family of Jefferson.[90] We accept as conclusive Park's acceptance of it as from the hand of Stuart. Jefferson's daughter, Martha Jefferson Randolph, is quoted by her granddaughter, Mrs. William B. Harrison, as saying she "thought it the best of the two hanging at Monticello.[91] Possibly Martha Jefferson Randolph was also the ultimate source of the statement by another of her granddaughters, Sarah Nicholas Randolph, that "this is the portrait of him which best gives the shape of his magnificent head and its peculiar pose."[92]

Two episodes in the relations of Jefferson and Stuart have escaped mention in the discussion of the portraits. Stuart's improvidence was proverbial; in spite of his lucrative employment he was often embarrassed for money. A letter from G. Humbold to Jefferson, May 13, 1801, solicits aid in behalf of Stuart, as in distressing circumstances.[93] The only response we know was Jefferson's subsequent further commission to him. Somewhat later in date is the following letter to Stuart in Boston, evidently apropos of a request for some device, with the gift of a gold watch, still preserved, the case inscribed "Gilbert Stuart, Esq. with compliments of Thomas Jefferson, 1806":

---

[86] Neither a Jefferson nor a Madison seems to have remained in Stuart's studio at the time of his death in 1827. The Stuart benefit exhibition at the Boston Athenaeum in that year included the Doggett set of the five presidents. *Cf.* the works of Mrs. Swan already cited.

[87] Oil on soft wood panel, $26\frac{1}{2} \times 21\frac{3}{16}$ in. Information kindly supplied by Mr. William Sawitzky who has also corrected the provenance stated in Park's book.

[88] Samuel W. Price, *Old Masters of the Bluegrass*, Louisville, Ky., Filson, 1902, catalogues some of them, as Nos. 92, 260. E. A. Jones, in *Matthew Harris Jouett, Kentucky Portrait Painter*, Louisville, Speed Mem. Mus., 1938, publishes on p. 115 the painter's inventory after decease, dated July 1828, which includes the item "2 portraits of Thomas Jefferson one with a frame $50."

[89] Oil on canvas, 19 × 19 in., sight, circle of $17\frac{1}{4}$ in. Hart was led to question its identity with the life portrait of the type, owing to Jefferson's careless allusions to it when it was out of his hands, in 1813, as "in water colours," and in 1819, as "on paper with crayons." G. C. Mason, in his *Life and Works of Gilbert Stuart*, Scribner, 1879, wrote that the date 1804 was "on the picture"; we have seen that this date is slightly erroneous; it is not Stuart's, but follows family tradition.

[90] As the history of the painting has not been exactly stated by Park and others, we may set it down here. The picture, though reserved from sale, was sent to Boston in 1828 with the rest of Jefferson's collection, and included in the exhibition of that year at the Athenaeum as No. 314, "Thos. Jefferson, an original in profile, Gilbert Stuart." Joseph Coolidge wrote on August 7, 1828, to his mother-in-law, Mrs. Martha Jefferson Randolph: "The portraits of Mr. Jefferson and Mr. Madison have been brought to my own house, for safe keeping—when you have decided on a place of residence they shall be forwarded to you" (letter kindly communicated by the late Harold Jefferson Coolidge). It passed ultimately into possession of the Coolidges and descended to the present owner.

[91] Letter printed in *The History of the Centennial Celebration of the Inauguration of George Washington*, ed. by C. W. Bowen, 485, 1892, where the statement of Sarah Nicholas Randolph is also printed.

[92] "His head . . . is set rather forward on his shoulders; and his neck being long, there is, when he is walking or conversing, a habitual protrusion of it." Daniel Webster's Correspondence, quoted in H. S. Randall, *The Life of Thomas Jefferson* III; 858, 1858.

[93] Library of Congress, Jefferson Papers, at present inaccessible.

FIG. 18. Stuart's third life portrait of Jefferson, painted in 1805. Private collection.

Monticello, May 30, '06

Dear Sir

Please accept this mark of my esteem. May you wear it with the pleasure of knowing it is presented to you from a friend who appreciated your great talent and the success you had with such a very poor subject. A few sketches such as need not take you a moment, will greatly oblige me.

I salute you with friendship

TH: JEFFERSON [94]

That Stuart in his last years contemplated painting a full length of Jefferson from life has been presumed on the basis of a letter of Jefferson written January 19, 1825, of which a short excerpt was published by John Hill Morgan in his biographical sketch of Stuart.[95] This letter, however, was actually written to the sculptor William J. Coffee, who had previously stayed at Monticello, and it has nothing to do with Stuart.[96]

The portraits of Jefferson by Stuart, especially the two surviving life portraits, have the characteristic freshness of observation of his works, and are among the most notable and precious representations of the sitter.

Benjamin Henry Latrobe, the gifted architect whom Jefferson in 1803 made Surveyor of Public Buildings of the United States, made sketches from life of a number of celebrated characters. Among these is a head of Jefferson, delicately drawn in pencil (fig. 19).[97] Latrobe had come from England to Virginia in 1796, to Philadelphia in 1798. To judge by the portrait itself, which shows locks of hair falling on the forehead, it comes from Jefferson's time as President, when Latrobe saw him familiarly from 1802 onward. We have no statement by Latrobe himself or by his heirs as to the circumstances under which this drawing was made. It does not conform to any of the surviving painted portraits of Jefferson, and we must presume it was taken by Latrobe from life. It would have been characteristic of Latrobe to sketch Jefferson, as he did Washington and others, early in their acquaintance, so that we may date the portrait about 1802.

Two small engravings of Jefferson, of closely related type, were made by Saint-Mémin, the gifted exile whose custom it was to make the profile of his sitters by his "physiognotrace," fill in the details with crayon and supply the sitter, for twenty-five dollars, with the original and twelve small engravings. According to Saint-Mémin's own notes on his engraved set at the Corcoran Gallery, one of these Jefferson engravings, in an oval medallion, was made in 1804, the other, in a circular medallion, in 1805.[98] A crayon original belonged to George Bancroft, who had a number of important and authentic Jefferson relics, now the property of the Worcester Art Museum (fig. 20).[99] Actually there is very little difference in the likeness or details in the two engravings, and we may be permitted to doubt Norfleet's belief that Jefferson sat for Saint-Mémin more than once. That this was in 1804 is shown by an item, hitherto unobserved, in Jefferson's accounts, on November 27, 1804: "Gave St. Menin order on bk. U. S. for 29.50," which at the usual rates covered three dozen extra impressions. In an advertisement with the date February 25, 1805, Saint-Mémin states that he "has on hand a few likenesses of the President of the United States, engraved by himself." These may well have been exhausted by popular demand, which thus led to the engraving of a second plate.

In 1805, Rembrandt Peale went with his father to Washington. He was to paint the portraits of celebrities for the Museum and, if possible, obtain other commissions. The following are quotations from unpublished letters of Charles Willson Peale:[100]

---

[94] Facsimile kindly communicated by the owner of the watch, Mr. James M. Swartz of Baltimore, who has also supplied a facsimile of a letter of Stuart: "Boston, May 6, 1809. Dear Mr. Denby: I acknowledge to have received the loan from you of one hundred dollars on the President's gift to me. I shall discharge it promptly. G. Stuart." There is an account of the matter in the *Baltimore American* for June 18, 1933.

[95] From a photostat furnished Mr. Morgan by Mr. Harry MacNeill Bland, who kindly also supplied us with one, the salutation being merely "Dear Sir."

[96] It is preserved by the Massachusetts Historical Society. Mr. Allyn B. Forbes, librarian of the Society, has kindly verified that it is addressed to Coffee.

[97] Reproduced from the drawing now in the possession of Miss Frances C. Semmes and reproduced through her kind permission and the courtesy of the Maryland Historical Society where it is on loan. Ferdinand C. Latrobe, great-grandson of the architect, believes the portrait not to be by Benjamin H. Latrobe, but perhaps by his son, John H. B. Latrobe, born 1803. If so it can only be copied after a lost drawing by his father. The paper on which it is drawn is without water mark as James W. Foster, Director of the Maryland Historical Society, kindly informs me.

[98] F. Norfleet, *Saint-Mémin in Virginia*, 30–31, 176–177, Richmond, Dietz, 1942, where the two documents are assembled. The two engravings are reproduced on p. 111. Norfleet supposed that Jefferson sat for Saint-Mémin early in 1804 and again in 1805.

[99] Crayon on paper, $23\frac{7}{8} \times 17\frac{5}{8}$ in.

[100] Generously supplied by Mr. Charles Coleman Sellers.

FIG. 19. Jefferson, pencil drawing, by Benjamin Henry Latrobe, about 1802. Maryland Historical Society.

FIG. 20. Jefferson, crayon drawing by Saint Mémin, 1804. Worcester Art Museum. Photograph by courtesy of the National Gallery of Art.

Jan. 19, 1805.

This morning we visited the President and obtained his consent to set for his Portrait, which Rembrandt will begin next Tuesday. Tomorrow he finishes Mr. Finley's picture which is a charming portrait. What other Portraits we shall make for the Museum I cannot yet determine.

Jan. 24.

Mr. Jefferson sat yesterday & today the 2nd setting. It is a charming portrait.

Jan. 30.

We wait on the President tomorrow morning to retouch his Portrait. Rembrandt has had another stye which however has passed off better than the other, lasting only two days. He will be able to paint tomorrow.

Postscript Feb. 1st.

The Portrait of Mr. Jefferson was finished yesterday, and much to our satisfaction.

No time was lost in turning the portrait to account, for on the very day it was finished the following advertisement appeared in the *National Intelligencer*:[101]

Those Gentlemen who are desirous of obtaining copies of the portrait of the President, just finished, by Rembrandt Peale, are requested to call at his painting room at Miss Dashfield's, Pennsylvania Avenue, where the picture can be seen for a few days during which Mr. Rembrandt Peale will remain in this City.[102]

We know of no such orders at this time. Mr. Sellers writes us:

Charles Willson Peale left for home February 3, leaving Rembrandt engaged to paint Gallatin and considering an historical picture of the trial of Samuel Chase. Charles Willson Peale feared his son's price of $50 for a portrait was too high, but hoped he might get some miniatures. Rembrandt reached Philadelphia on February 15, bringing a "charming" portrait of Gallatin, and his father felt satisfied that his improvement in his art had more than repaid them for the trip to Washington. Apparently the financial return from their paintings and polygraphs had been just about nil.

As we have noted, the catalogue of the sale of the Peale Museum in 1854 includes the item

152. Thomas Jefferson, by Rembrandt Peale.

[101] Published by J. H. Morgan in *Virginia Historical Portraiture*, 252, 1930.

[102] Generously supplied by Mr. Charles Coleman Sellers.

This is believed with much reason to be the portrait[103] given by Thomas J. Bryan to the New-York Historical Society in 1867 (fig. 21). It is a bust, three-quarters to left, showing the sitter in a fur-lined coat above which shows a red neckcloth about the white stock. The hair, now graying, is arranged as in Stuart's portrait of the same year.

Another Rembrandt Peale Jefferson, identical in type (fig. 22),[104] was owned in 1896 by William H. Perkins of Baltimore, who stated at that time that it was "painted from life for Captain Muller (or Mullen), of Germantown, Pa.; purchased from his widow in 1849 or 1850, by a Baltimore family; obtained from them by William H. Perkins." It subsequently belonged to the late Harold Jefferson Coolidge and is now owned by his son Laurence Coolidge.[105]

From personal examination I am convinced that the New-York Historical Society portrait is without doubt the one taken from life, and that the other is a replica. In the New York picture there is substantially more literal detail, especially about the eyes. The face is sterner, more realistic, whereas the Coolidge portrait, though very faithfully repeated, is softer and more idealized. At sixty-two Jefferson is beginning to show his age, though the fire of the eyes is undiminished. Along with the pictures by Mather Brown, Trumbull, Charles Willson Peale, Stuart and, later, Sully, the life portrait by Rembrandt Peale is among the admirable and worthy characterizations of the sitter, and one of the finest works of the artist.

Joseph Delaplaine, desiring a portrait of Jefferson to be engraved for his published Portrait Gallery, ultimately proposed to send Bass Otis to take one from life. Jefferson replied on May 20, 1816, that he would be happy to receive him.[106] "With respect to the merit of Otis's painting," Jefferson wrote diplomatically to William Thornton, "I am not qualified to say anything, for this is a case where the precept of 'Know thyself'

[103] Oil on canvas, 28 × 23½ in.

[104] Oil on canvas, 29 × 24 in.

[105] While the family was aware that it was acquired, not by descent but by purchase, the memory of its source had been lost. Its identity is established beyond all doubt by the coincidence of every detail of the *craquelure* with an old photograph of the Mullen-Perkins painting acquired by the Frick Art Reference Library from the collection of Charles Henry Hart, along with a letter of William H. Perkins, then the owner, dated October 28, 1896, supplying the information as to its origin as given above.

[106] Library of Congress, Jefferson Papers.

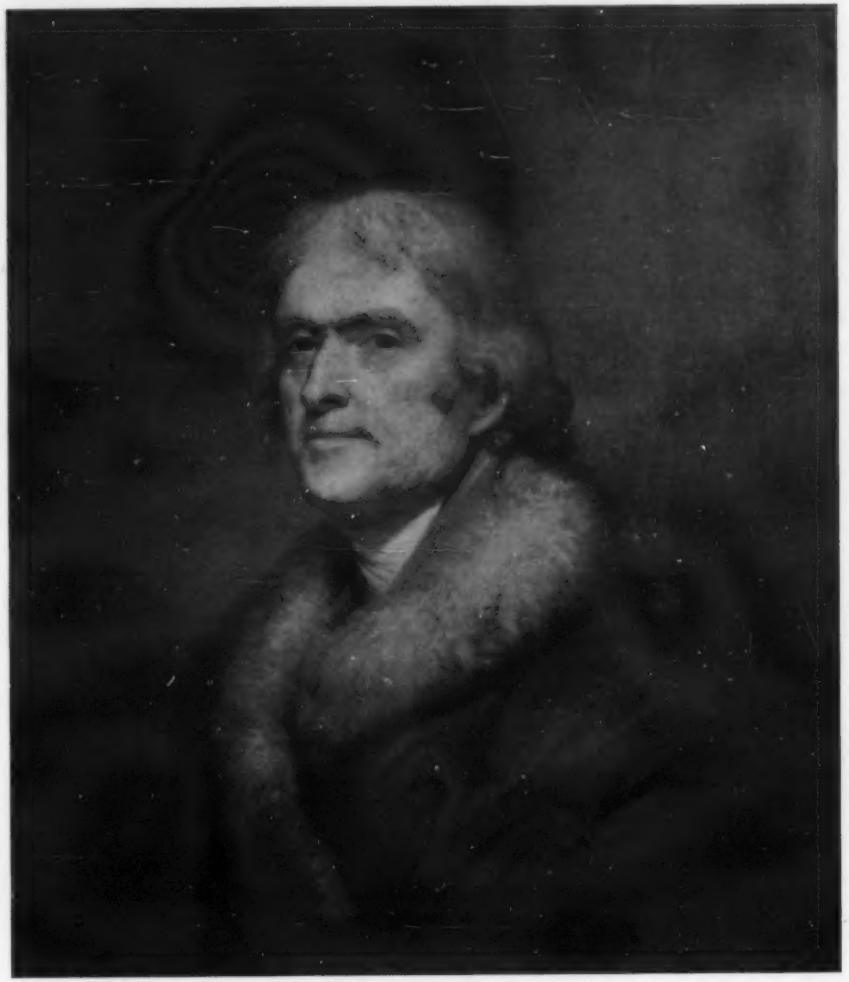


FIG. 21. Jefferson, life portrait by Rembrandt Peale, 1805. Courtesy of The New-York Historical Society.

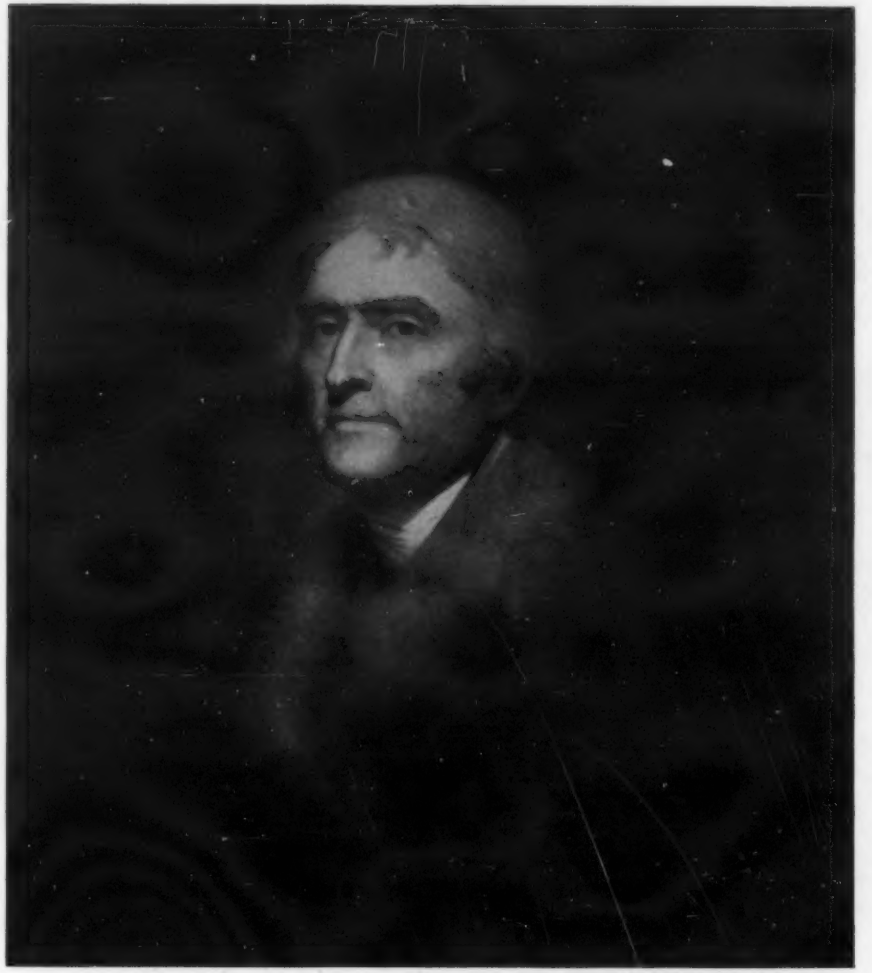


FIG. 22. Jefferson, replica by Rembrandt Peale of his life portrait. Lieutenant Lawrence Coolidge, U.S.N.R. Photograph by courtesy of The National Gallery of Art.

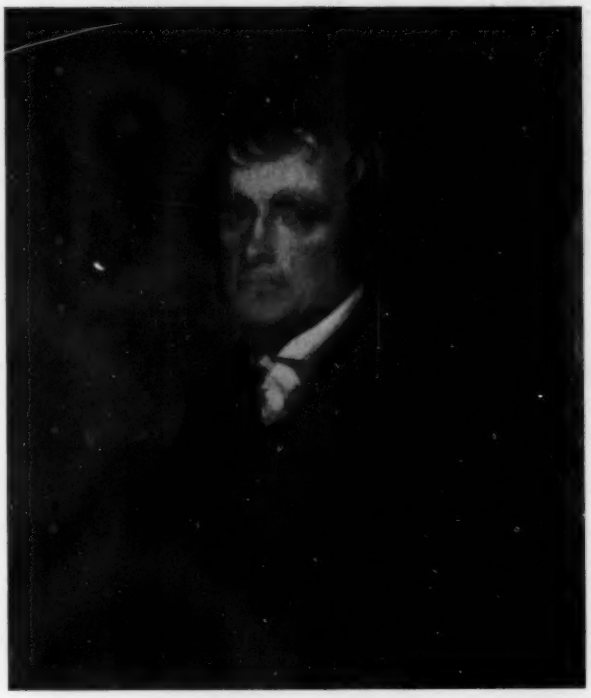


FIG. 23. Jefferson, by Bass Otis, 1816. Thomas Jefferson Memorial Foundation, Monticello, Virginia. Photograph by courtesy of The National Gallery of Art.

does not apply."[107] Thornton had been less restrained in his expression, writing:[108] "Never was such injustice done to you except by sign painters and General Kosciusko, than which last nothing can be so bad, and when I saw it I did not wonder he lost Poland—not that it is necessary for a general to be a painter, but he should be a man of such sense as to discover that he is not a painter.[109]

The engraving by John Neagle published in Delaplaine's *Repository*, 1817, shows the hands folded in the lap. Several Jefferson portraits by or attributed to Bass Otis exist. A standing half-length, the left hand holding a book, owned in 1892 by the late William J. Campbell,[110] was, as the only one then known, naturally supposed by Mr. Hart to be the life portrait. Numerous others have since appeared. One, without hands, included in the exhibition of the National Gallery (fig. 23),[111] was presented in 1930 to the Jefferson Memorial Foundation by Mrs. E. M. Bolles, granddaughter of John C. Trautwine, the celebrated engineer. On the back of the canvas is the inscription: "Painted from life by my father, Bass Otis; and given by me to John C. Trautwine." At the exhibition of historical portraits held at the Pennsylvania Academy of the Fine Arts in 1926, no. 104, catalogued as a Bass Otis Jefferson, was lent by Miss M. E. McManus of Merion, Pennsylvania. What purported to be this picture was illustrated as no. 52 in the sale of Hiram Burlingham at the American Art Association–Anderson Galleries, on January 11, 1934.[112] This was acquired by Colonel David K. E. Bruce of Staunton Hill, Virginia. It is a full bust, without hands, within a painted oval.

We do not undertake to determine whether any one among these rather pedestrian works is actually a life portrait.

Jefferson was requested by Jared Mansfield in 1821 to give sittings to Thomas Sully for a portrait to be presented to the Military Academy at West Point. He replied with appreciation and humor, saying:

> Mr. Sully, I fear . . . will consider the trouble of his journey and the employment of his fine pencil, as illy bestowed on an ottamy of 78. Voltaire when requested by a female friend to sit for his bust by the sculptor Pigalle, answered, "J'ai soixante seize ans; et M. Pigalle doit, dit-on, venir modeler mon visage. Mais, Madame, il faudrait que j'eusse un visage. On n'en devinerait à peine la place: mes yeux sont enfoncés de trois pouces; mes joues sont de vieux parchemin mal collés sur des os qui ne tiennent à rien." I will conclude, however, that what remains is at your service and that of the pencil of Mr. Sully.[113]

---

[107] July 27, 1816, *The Writings of Thomas Jefferson*, ed. by A. E. Bergh and A. A. Lipscomb, XVIII: 296–297, 1907.

[108] July 20, 1816. Jefferson Papers, Library of Congress, quoted by Hart in *McClures Magazine*.

[109] The Kosciusko profile is known only from engravings, and could, moreover, not have been from life, as it represents Jefferson at an age far greater than on their last meeting.

[110] Mr. William J. Campbell, Jr., informs me this was sold "about 30 years ago" to Charles F. Gunther of Chicago, whose collection was subsequently sold at auction "in New York." In the sale by Stan. V. Henkels at Philadelphia, November 13, 1917, of "Historical Oil Portraits from the collection of Charles F. Gunther, of Chicago and from other sources," no. 26 is a Jefferson "Full bust. Size 25 × 30. Canvas," listed as of unknown authorship and not illustrated, which may be this picture. The account of the sale in the *Art News* at that time does not list the purchaser.

[111] Oil on canvas, 30 × 25 in.

[112] Oil on panel, 14 × 12 in., head and shoulders, with unbuttoned coat, within a painted oval.

[113] February 13, 1821. *The Writings of Thomas Jefferson*, ed. by A. E. Bergh and A. A. Lipscomb XV; 314, 1907.

FIG. 24. Life portrait "From Jefferson, 1821; finished 1830, T. S.," by Thomas Sully. American Philosophical Society.

Sully came to Monticello in March and painted the bust portrait now belonging to the American Philosophical Society (fig. 24).[114] It is signed on the back with Sully's cipher, and the inscription: "From Jefferson 1821. Finished 1830." On June 18 of the latter year, it was given to the Society by Jefferson's friend and admirer William Short, the price as recorded in Sully's register being $75.00.[115] There is no question that this is the portrait painted from life; just above the head we see the trace of the artist's revisions. It is a brilliant and living portrait, soundly painted and thus still in very fine condition, free from repaint, so that we can depend on the fidelity of the limpid coloring. Jefferson's eyes are a grayish blue-green; the hair and eyebrows are still not wholly white, but show traces of their sandy hue.[116]

Mr. John Hill Morgan lent to the exhibition in Washington a small study (fig. 25),[117] not listed in Sully's register, for the West Point portrait, the history of which is stated as follows:

> This and other Sully sketches [were] bequeathed by Thomas Sully in 1872 to Blanche Sully; she bequeathed them to Thomas Francis Sully Darley; he bequeathed them to Thomas Nash, New York architect, from whom they were bought by Mrs. A. W. Sully and André Reuff, Brooklyn. This sketch of Jefferson was acquired from Mrs. Sully and Mr. Reuff by Mr. John Hill Morgan.[118]

In this sketch the idea evidently was to suggest Jefferson's taking the oath of office as President. He stands in the portico of a building, at the head of a broad flight of steps beside a Tuscan column. Behind him are two other personages, one perhaps a judge; in the background to the right, before a wing of the building, is a multitude of figures—seemingly a suggestion of an inauguration. The building is neither the Capitol nor Monticello, yet it could readily have been sketched, with free modifications at the east portico of Jefferson's house, the figure from life.

The second Jefferson entry in Sully's register is a small full-length study for the West Point picture, "begun March 27th, 1822, finished April 10, 1822. Size 29 × 18 in. Price $150.00." This is apparently the painting now owned by Mr. Edward S. Moore of New York (fig. 26)[119] said to have been "sent to be sold at Earl's Gallery in Philadelphia in 1830 . . . there purchased by a Mr. Hammond of Columbia . . . remained in his family for four generations."[120] A. S. Salley, Secretary of the Historical Commission of South Carolina, informs me that it descended from Governor Hammond to his son, E. S. Hammond, and then to the present James H. Hammond; prior to its sale it being for some years in the Capitol at Columbia. This study fully forecasts the treatment of the large final picture, as to composition, pose, and treatment. The idea of showing Jefferson on an exterior portico was abandoned and for the Tuscan column was substituted one of the Corinthian order, in a breccia marble with a row of leaves about the base. Column and base alike were those of the new House of Representatives (now Statuary Hall), then lately put in use.

The large full-length at West Point (fig. 27),[121] according to Sully's register, was begun April 12, 1822, and finished May 7, 1822, the price being $500. It bears Sully's cipher and the date 1822. This noble work, with the sketches for it, is, except for the figure of Jefferson in Trumbull's Declaration of Independence, our only graphic evidence of his stature, as witnessed by an artist who painted him from life.

---

[114] Oil on canvas, 30 × 25 in. The picture has been relined. The inscription has been copied in facsimile on the new lining.

[115] E. Biddle and M. Fielding, *The Life and Works of Thomas Sully*, Phila., 1921. This lists as nos. 914–918 five Jeffersons by Sully, including two late replicas, beside mentioning the replica or copy once owned by James Monroe and now by the Jefferson Society of the University of Virginia, which is not mentioned in Sully's register.

[116] His head is still well covered with hair which, having been once red, and now turning gray, is of an indistinct sandy color." Webster in 1824, quoted in Randall's *The Life of Thomas Jefferson* III: 504, 1858. Surviving locks of Jefferson's hair taken by members of his family after his death still show very distinct color.

[117] Oil on cardboard, 9¾ × 6⅞ in. Sully frequently made such oil sketches on card.

[118] Oral information by André Reuff, May 12, 1932, to the Frick Art Reference Library. Mr. Henry W. Kent, a friend of the late Thomas Nash, kindly informs me that he knew the material while in his hands, as well as the various other parties to the transaction.

[119] Oil on canvas, 29⅛ × 18 in. Relined with the following inscription, apparently copied in facsimile, on the back of the new canvas: "Thomas Jefferson/Ex President of the United States/Painted by T. Sully/1821/TS."

[120] Helen Comstock, Jefferson in His Old Age, *International Studio* **96**: 17–18, 1930, where it is reproduced in color.

[121] Oil on canvas, 103 × 67 in., photograph kindly supplied by Lt. Col. John P. Maher.

FIG. 25. Jefferson, sketch by Thomas Sully for the full length. John Hill Morgan, Esq. Photograph by courtesy of the National Gallery of Art.

FIG. 26. Jefferson, by Thomas Sully, 1822, the small full length. Edward S. Moore, Esq. Photograph by courtesy of the National Gallery of Art.

FIG. 27. Jefferson, by Thomas Sully, the large full length, 1822. United States Military Academy.

Another artist who portrayed Jefferson from life was the sculptor William J. Coffee, who later made for Jefferson the ornaments of the cornices of the University and at Poplar Forest. Jefferson wrote to Madison April 11, 1821, commending Coffee as "really able in his art," and saying he had given "less trouble than any artist, painter, or sculptor I have ever submitted myself to." The next day, Jefferson noted sending "a draft . . . for 105 D. in favor of Mr. Coffee, a sculptor. This is to pay for the originals of 3 busts, to wit Mrs. Randolph's, Ellen's and mine. I bespoke of him 12 copies of Mrs. R.'s @ 5 D. each." A further payment of $65 to Coffee for busts was made March 18, 1820, with a reference to this earlier entry, so that we may presume these were the twelve copies of Mrs. Randolph's bust. Coffee returned to Monticello in 1822, when on March 25 Jefferson paid him $100 for restoring paintings of his collection. Shortly before this, on February 7, there is an item of $50 "inclosed to Ellen Randolph for busts." Several such busts in terra cotta, less than life size, presumably made by Coffee, are known to survive. One is now at Monticello, having been presented by a Jefferson descendant as a portrait of Anne Cary Randolph Bankhead, Jefferson's eldest grandaughter. Others in the family have passed as being of others of the Randolph sisters. They are of sufficient merit to make us regret the loss of Coffee's bust of Jefferson.

On January 1, 1825, Coffee wrote to Jefferson from New York that he was shipping some cast ornaments to the University (of the construction of which, under Jefferson, A. S. Brockenbrough had charge), and asked him to sit for a statue.[122] To this letter Jefferson replied:[123]

Monticello, Jan. 19. '25.

Dear Sir:

Your favor of the 1st came to hand on the 7th. I immediately sent to Mr. Brockenbrought the one directed to him. . . . I have for some time refused any more to set for portraits or busts. But unwilling to refuse you, and your proposition being for a whole length, takes it out of my rule. So I shall willingly comply with your request on that subject. You know that here you are a welcome guest always and expected always to make it your headquarters.

Ever with esteem and respect yours

TH. JEFFERSON

[122] Library of Congress, Jefferson Papers.

[123] Massachusetts Historical Society, Jefferson Papers. This is the letter which has been mistakenly published, in part, as to Stuart.

Coffee, however, did not leave New York that year, and his project for a statue never came to execution.

We shall not rehearse the oft-repeated story[124] of the taking of the life-mask of Jefferson, the year before his death, by John H. I. Browere. Jefferson's letter to Madison, October 18, 1825,[125] sufficiently details the hardships and even dangers of the process. The resulting plaster half-length—for such it is, rather than a mask—survives, though more than once repaired, the property of the New York State Historical Association at Cooperstown, New York (fig. 28). At eighty-two, in spite of the inevitable decay of age,[126] the fire and determination of Jefferson were still alive. It was after the ordeal of its making that Jefferson wrote: "I now bid adieu forever to busts and even portraits."

Jefferson himself was guarded in any expression of opinion as to the merit of his portraits. On May 30, 1813, he wrote to the publisher, Joseph Delaplaine, as to

> . . . whether there is an approved portrait of myself . . . Mr. Stuart has drawn two portraits of me, at different sittings of which he prefers the last. He also drew a third in water colours, a profile in the medallion stile. Mr. Rembrandt Peale also drew a portrait in oil colours on canvas while I lived in Washington. Of the merit of these I am not a judge, there being nothing to which a man is so incompetent a judge as his own likeness. He can see himself only by reflection, and that of necessity full-face or nearly so.[127]

Again in 1816 he wrote to Delaplaine:

> You must judge for yourself whether a copy by Stewart himself . . . whose name is without rival

[124] C. H. Hart, Unknown Life Masks of Great Americans, *McClure's Magazine* 9, 1897; *Browere's Life Masks of Great Americans*, Doubleday, 1899. *Life Masks of Noted Americans*, exhibition under the sponsorship of the New York State Historical Association, 1940.

[125] *The Writings of Thomas Jefferson*, ed. by A. E. Bergh and A. A. Lipscomb, XIX: 287, 1907.

[126] Unlike Washington, however, Jefferson had kept his teeth. His grandson, Thomas Jefferson Randolph, wrote, "Dying in his 84th year, he had not lost a tooth, or had one defective." Quoted by H. S. Randall, *The Life of Thomas Jefferson* III: 674, 1858.

[127] Library of Congress, Jefferson Papers.

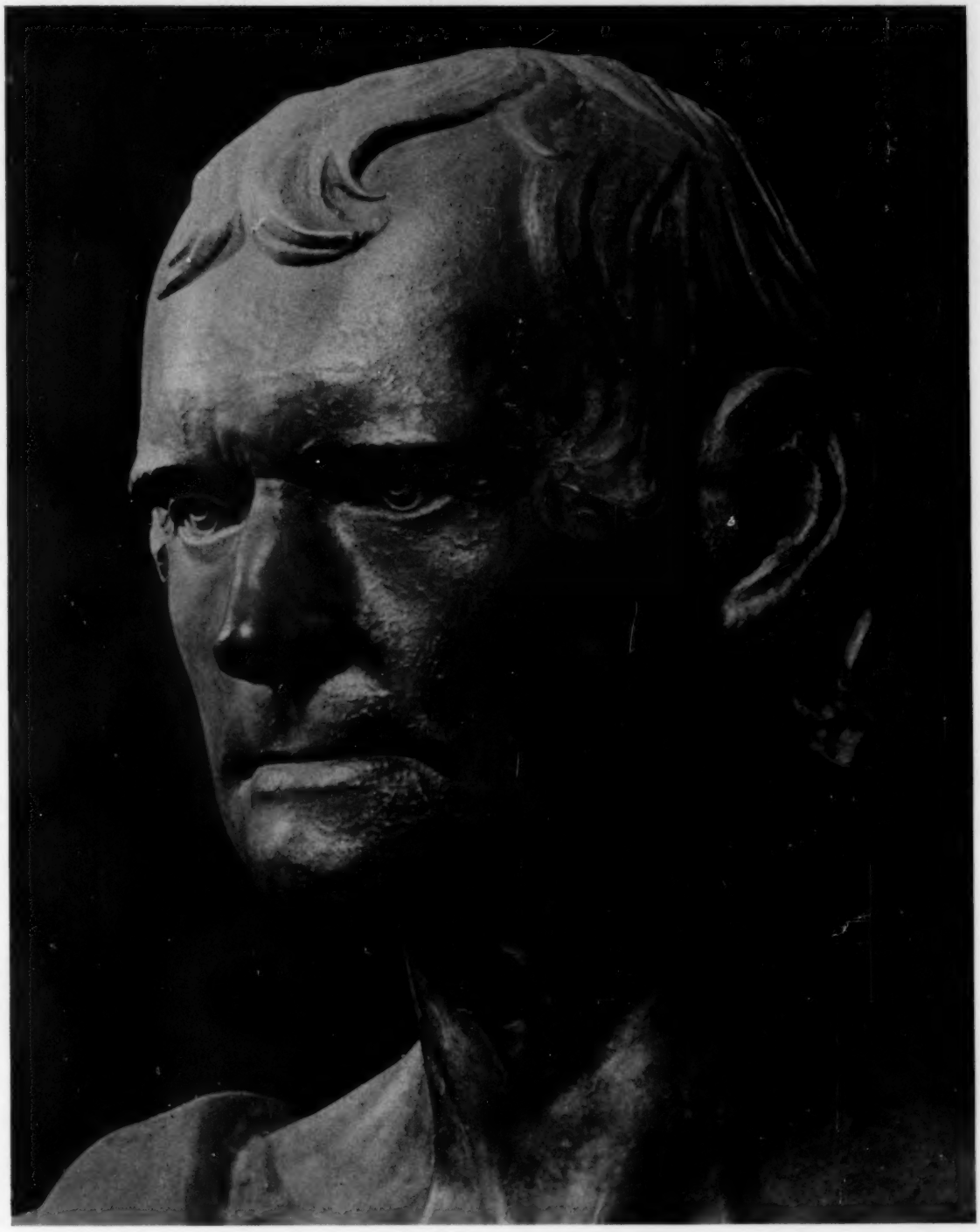


FIG. 28. Jefferson, original plaster life mask by Browere, 1825. New York State Historical Association.

would not be more valuable than an original by any other artist however good. . . .

In spite of the diplomatic caution which years of sad experience with quotations had enjoined, Jefferson here shows the same prophetic anticipation of future judgments for which he is celebrated in other fields. The Sully was still to be painted. Had the form of the question admitted, Jefferson would surely have included also the sculptured busts he admired, and the likeness in Trumbull's composition. It is the Jefferson of Stuart, of Rembrandt Peale and of Sully who is the Jefferson of posterity.

INDEX TO VOLUME 88

Agallian leafhoppers, Some viruses transmitted by (Black), 132
Aviation, military, Human problems in (Bronk), 189

Binaries, eclipsing, Some tables to facilitate computation of elements of (Kopal), 145
Black, L. M., Some viruses transmitted by agallian leafhoppers, 132
Bloch, Henry Simon, Fiscal policy and social reform as it may affect the potentialities of the personal income tax, 36
Blood, blood derivatives, and blood substitutes (Cohn), 159
Blough, Roy, Conflict and harmony in taxation, 22
Bronk, Detlev W., Human problems in military aviation, 189
Buehler, Alfred G., The sales tax, 29
Butler, Harold B., The International Labor Organization lessons for twenty-five years, 214

Capitalist and the social structure (Warren), 1
Castaways, The water requirements of (Gamble), 151
Catawba pottery-making (Fewkes), 69
Cherokee pottery-making (Fewkes), 69
Cohn, Edwin J., Blood, blood derivatives, and blood substitutes, 159
Coiling and pottery-making (Fewkes), 69
Corporations, Post-war program for taxation of (Seligman), 60

Dakota treatment of murderers (Deloria), 368
Davenport, Charles Benedict, Postnatal development of the human extremities, 375
Deloria, Ella Cara, Dakota treatment of murderers, 368
Diseases, infectious, Trends in the study and control of (Dubos), 208
Doriot, Georges F., Environmental protection, 196
Drought, Recovery of midwestern prairies from (Weaver), 125
Dubos, René J., Trends in the study and control of infectious diseases, 208
Dyer, R. E., The role of immunization in wartime advances in medicine, 182

Environmental protection (Doriot), 196

Fewkes, Vladimir J., Catawba pottery-making, with notes on Pamunkey pottery-making, Cherokee pottery-making, and coiling, 69
Figural after-effects, an investigation of visual processes (Köhler and Wallach), 269
Fiscal and social aims, The harmonization of (Hunter), 43

Gamble, James L., The water requirements of castaways, 151

Haney, John L., "Of the people, by the people, for the people," 359
Hobbs, William Herbert, Mountain growth, a study of the southwestern Pacific region, 221
Hunter, Merlin H., The harmonization of fiscal and social aims, 43

Immunization in wartime advances in medicine (Dyer), 182
Income, national, Taxes and (Kuznets), 10
Income tax, personal, Fiscal policy and social reform as it may affect the potentialities of (Bloch), 36
Infections, Penicillin in the treatment of (Keefer), 174
Infectious diseases, Trends in the study and control of (Dubos), 208
International Labor Organization lessons for twenty-five years (Butler), 214

Jefferson, Life portraits and their replicas (Kimball), 497

Keefer, Chester S., The present status of penicillin in the treatment of infections, 174
Kimball, Fiske, The life portraits of Jefferson and their replicas, 497
Köhler, Wolfgang, and Hans Wallach, Figural after-effects, an investigation of visual processes, 269
Kopal, Zdeněk, Some tables to facilitate computation of elements of eclipsing binaries, 145
Kuznets, Simon, Taxes and national income, 10

Lagrange's personality (Sarton), 457
Lincoln's Gettysburg Address (Haney), 359

Medicine, Wartime advances in, symposium, 151–213
Mountain growth, a study of the southwestern Pacific region (Hobbs), 221

Neuropsychiatry (Strecker), 204
Newcomer, Mabel, The federal, state, and local tax structure after the war, 50

Pacific region, southwestern (Hobbs), 221
Pamunkey pottery-making (Fewkes), 69
Penicillin in the treatment of infections (Keefer), 174
Penrose Memorial Lecture (Butler), 214
Postnatal development of the human extremities (Davenport), 375
Pottery-making, Catawba, Pamunkey, and Cherokee (Fewkes), 69
Prairies, midwestern, Recovery from drought (Weaver), 125

Sarton, George, Lagrange's personality, 457
Seligman, Eustace, A post-war program for the taxation of corporations and their stockholders, 60
Shultz, William J., Post-war social orientation of state tax systems, 55
Social and fiscal aims, The harmonization of (Hunter), 43
Social orientation, Post-war, of state tax systems (Shultz), 55
Social reform and fiscal policy as it may affect the potentialities of the personal income tax (Bloch), 36
Social structure, The capitalist and (Warren), 1
  Taxation and, symposium, 1–67
Stars, nearby, A photographic study of the proper motion of (van de Kamp), 372
Strecker, Edward A., Neuropsychiatry, 204
Symposia:
  Taxation and the social structure, 1–67
  Wartime advances in medicine, 151–213

Tax, personal income, Fiscal policy and social reform as it may affect the potentialities of (Bloch), 36
Sales (Buehler), 29
structure, Federal, state, and local, after the war (Newcomer), 50
systems, state, Post-war social orientation of (Shultz), 55
Taxation, Conflict and harmony in (Blough), 22
and the social structure, symposium, 1–67
of corporations and their stockholders, Post-war program for (Seligman), 60
Taxes and national income (Kuznets), 10

Van de Kamp, Peter, A photographic study of the proper motion of nearby stars, 372
Viruses transmitted by agallian leafhoppers (Black), 132
Visual processes, Figural after-effects, an investigation of (Köhler and Wallach), 269

Wallach, Hans (with Wolfgang Köhler, see Köhler)
Warren, Robert, The capitalist and the social structure, 1
Water requirements of castaways (Gamble), 151
Weaver, John E., Recovery of midwestern prairies from drought, 125
Whipple, Allen O., Recent advances in the treatment of wounds, 177
Wounds, Recent advances in the treatment of (Whipple), 177

PUBLICATIONS

OF

The American Philosophical Society

The publications of the American Philosophical Society consist of PROCEEDINGS, TRANSACTIONS, MEMOIRS, and YEAR BOOK.

The PROCEEDINGS contains original papers which have been read before the Society in addition to other papers which have been accepted for publication by the Committee on Publications. Beginning with Volume 88 one volume will be issued each year and the price is $5.00 net per volume.

The TRANSACTIONS, the oldest scientific journal in America, was started in 1769; its quarto size makes it especially suitable for the publication of treatises and monographs requiring large illustrations. Beginning with Volume 34 one volume will be issued each year and the price is $6.00 net per volume. The annual subscription price for both the PROCEEDINGS and the TRANSACTIONS is $10.00.

Each volume of the MEMOIRS is published in book form. The titles cover a variety of subjects. The price of each volume is determined by its size and character; the prices have varied from $1.50 to $7.50.

The YEAR BOOK is of considerable interest to scholars because of the reports on grants for research and also because of the section devoted to the Library which contains in greater or less detail description of important holdings and accessions. In addition it contains also the Charter and Laws, and lists of Officers, Committees, and Members. The YEAR BOOK is published as soon as possible after the close of each calendar year. The price is $1.50.

A catalogue of the contents of the publications may be had
by addressing the

AMERICAN PHILOSOPHICAL SOCIETY

INDEPENDENCE SQUARE

PHILADELPHIA 6, PA.

THE AMERICAN PHILOSOPHICAL SOCIETY

Recent Publications

PROCEEDINGS: (The following are symposia.)

The Early History of Science and Learning in America.
Vol. 86, no. 1, 204 pp., 25 fig., 1942; Vol. 87, no. 1, 119 pp., 55 fig., 1943.

Post-War Problems.
Vol. 87, no. 2, 78 pp., 4 fig., 1943.

Thomas Jefferson.
Vol. 87, no. 3, 91 pp., 7 pl., 1943.

The Organization, Direction, and Support of Research.
Vol. 87, no. 4, 74 pp., 1944.

Taxation and the Social Structure.
Vol. 88, no. 1, 67 pp., 1944.

Wartime Advances in Medicine.
Vol. 88, no. 3, 69 pp., 6 fig., 1944.

TRANSACTIONS:

John Bartram: Diary of a Journey through the Carolinas, Georgia, and Florida, 1765–66.

William Bartram: Travels in Georgia and Florida, 1773–74; A Report to Dr. John Fothergill. Annotated by Francis Harper.
Vol. 33, pts. 1 and 2, 242 pp., 50 pl., 1942–1943. Either part paper-bound, $2.00; together cloth-bound, $5.00.

Infectious Anemias Due to Bartonella and Related Red Cell Parasites, by David Weinman.
Vol. 33, pt. 3, 108 pp., 1 pl., 1 map, 1944. $1.25.

The Velocity of Light, by N. Ernest Dorsey.
Vol. 34, pt. 1, 110 pp., 20 fig. $2.25.

The Arc Spectrum of Iron (Fe I), by Henry Norris Russell, Charlotte E. Moore, and Dorothy W. Weeks.
Vol. 34, pt. 2, 97 pp., 1944. $2.25.

MEMOIRS:

La Tradition Littéraire des Idéologues, by Emile Cailliet.
Vol. 19, 322 pp., 1943. $4.00.

Edward Tyson (1650–1708) and the Rise of Human and Comparative Anatomy in England, by M. F. Ashley Montagu.
Vol. 20, 488 pp., 57 illus., 1943. $5.00.

Sumerian Mythology: A Study of Spiritual and Literary Achievement in the Third Millennium B. C., by S. N. Kramer.
Vol. 21, 125 pp., 21 pl., 3 fig., 1944. $2.00.

Thomas Jefferson's Garden Book, annotated by Edwin M. Betts.
Vol. 22, 704 pp., 37 pls., 1944. $5.00.